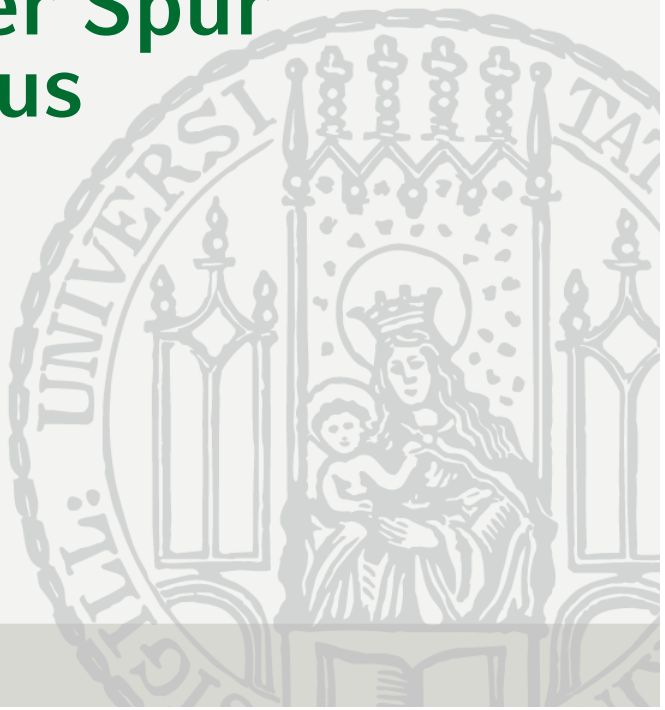
LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Prof. Karsten Suhre

Mit dem Computer auf der Spur von Mimi, dem Riesenvirus

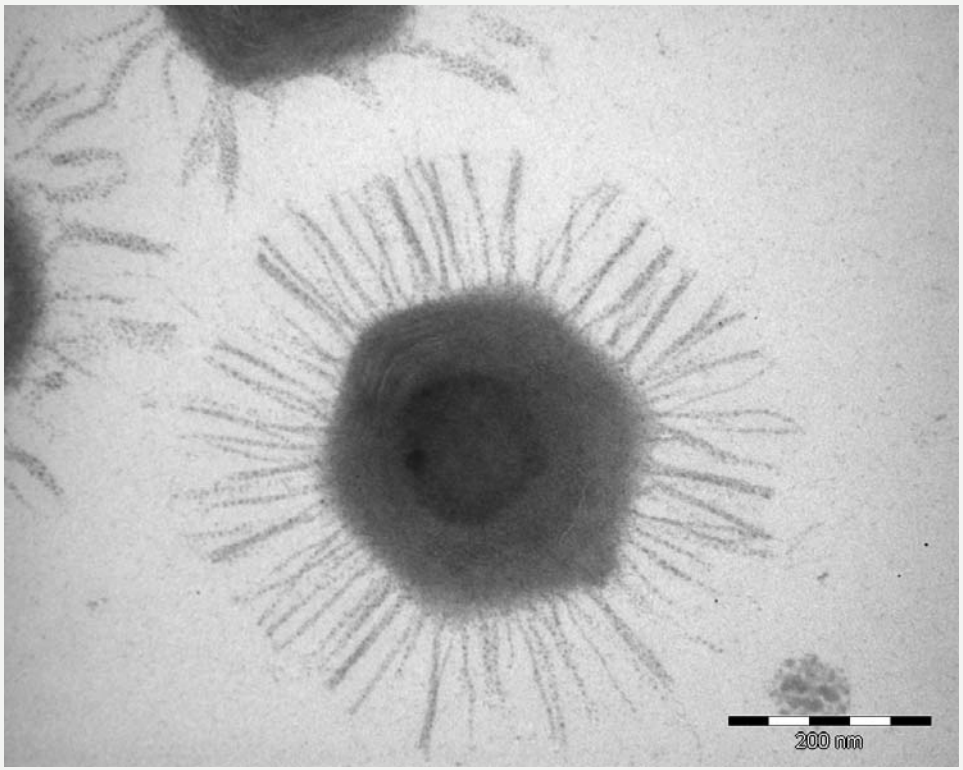
Seniorenstudium
Naturwissenschaften: Biologie
09.07.2007, 17:15 -18:00 Uhr



Worum geht es heute?

- Ein Riesenvirus wirft Fragen nach dem Ursprung der Viren auf.
- Wie die Bioinformatik dabei hilft, diese Fragen zu beantworten.

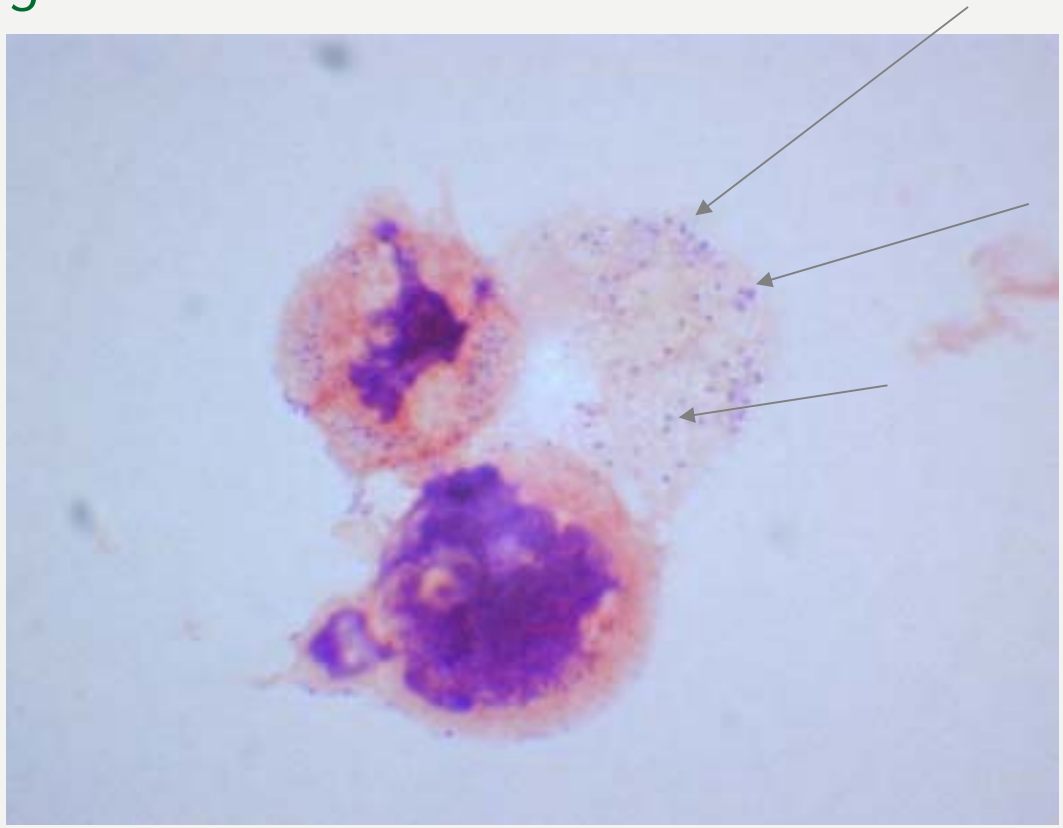
Das Riesenvirus *Acanthamoeba polyphaga mimivirus*



(c) Didier Raoult

Entdeckung des Mimivirus in Bradford, England, durch Dr. Tom Rowbotham, 1992.

Aber ... Mimivirus wurde zunächst für ein Bakterium gehalten → *Bradfordcoccus*



(c) Didier Raoult

Nach 10 Jahren vergeblicher Versuche, den *Bradfordcoccus* zu kultivieren, kommt er nach Marseille ...

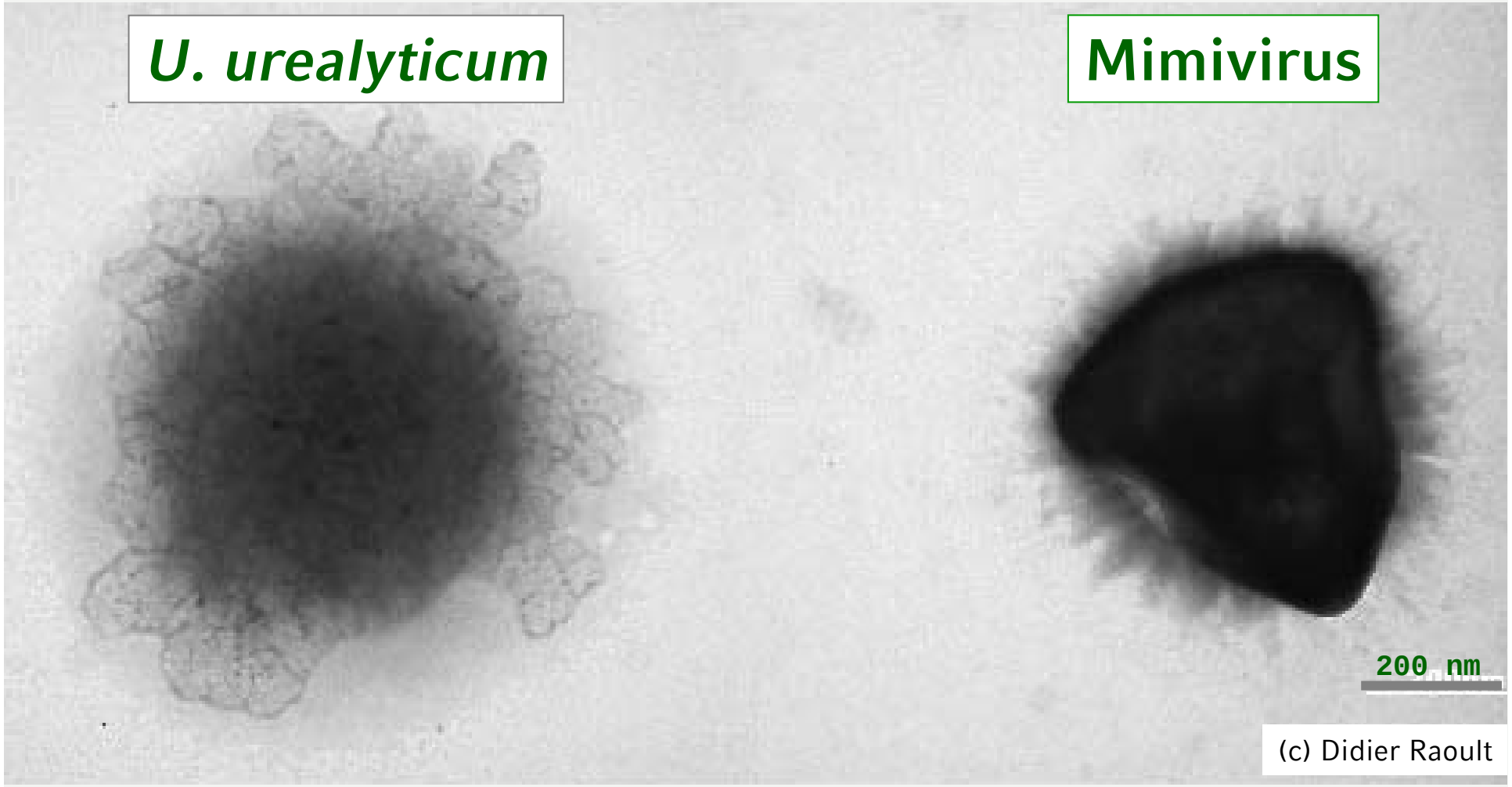
Didier Raoult
CNRS Marseille
Kliniker

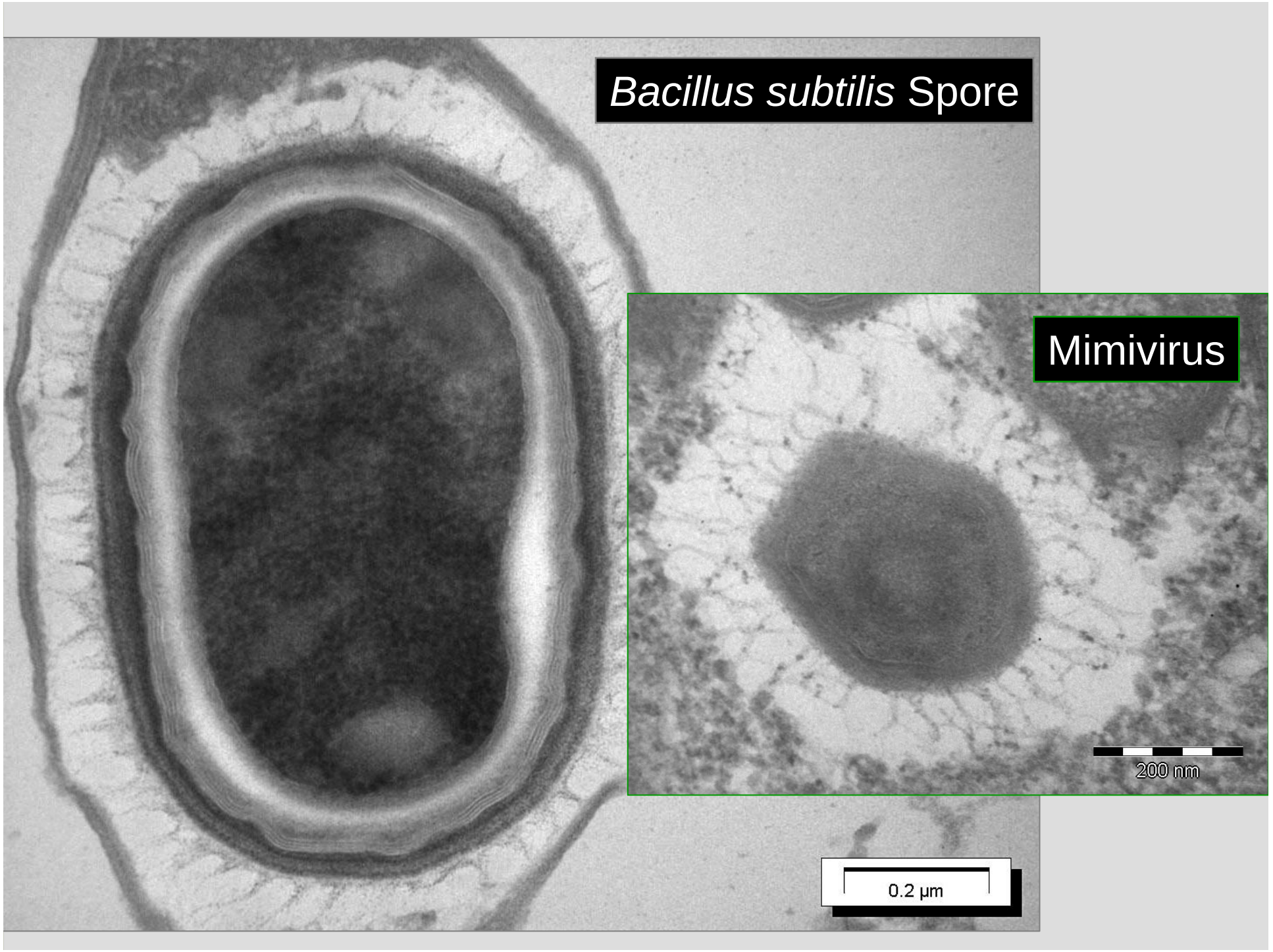
Jean-Michel Claverie
CNRS Marseille
Bioinformatiker

Die Überraschung ....

... es handelt sich um ein Virus!

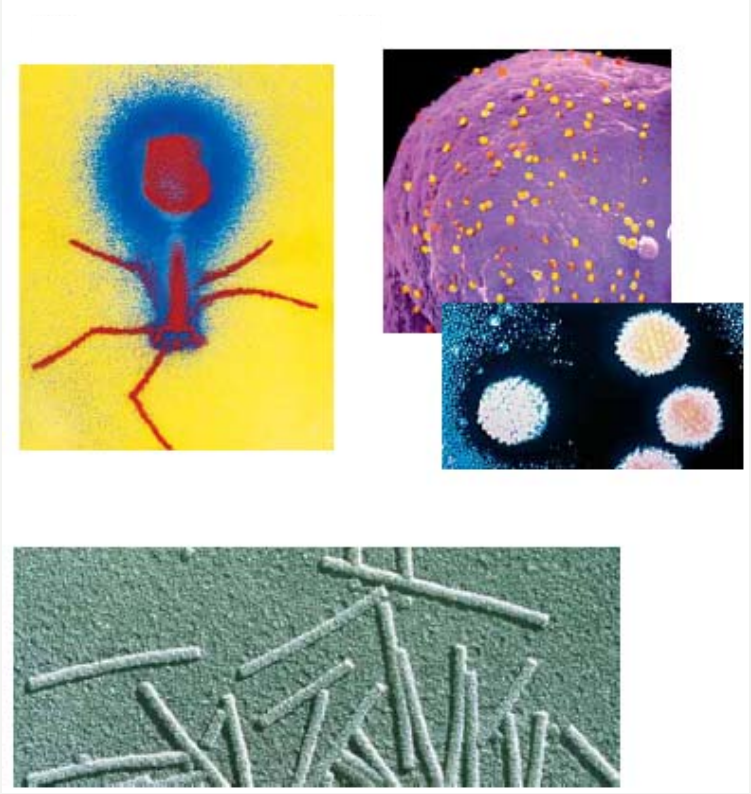
... so groß wie ein Bakterium!


U. urealyticum
Mimivirus
200 nm
(c) Didier Raoult

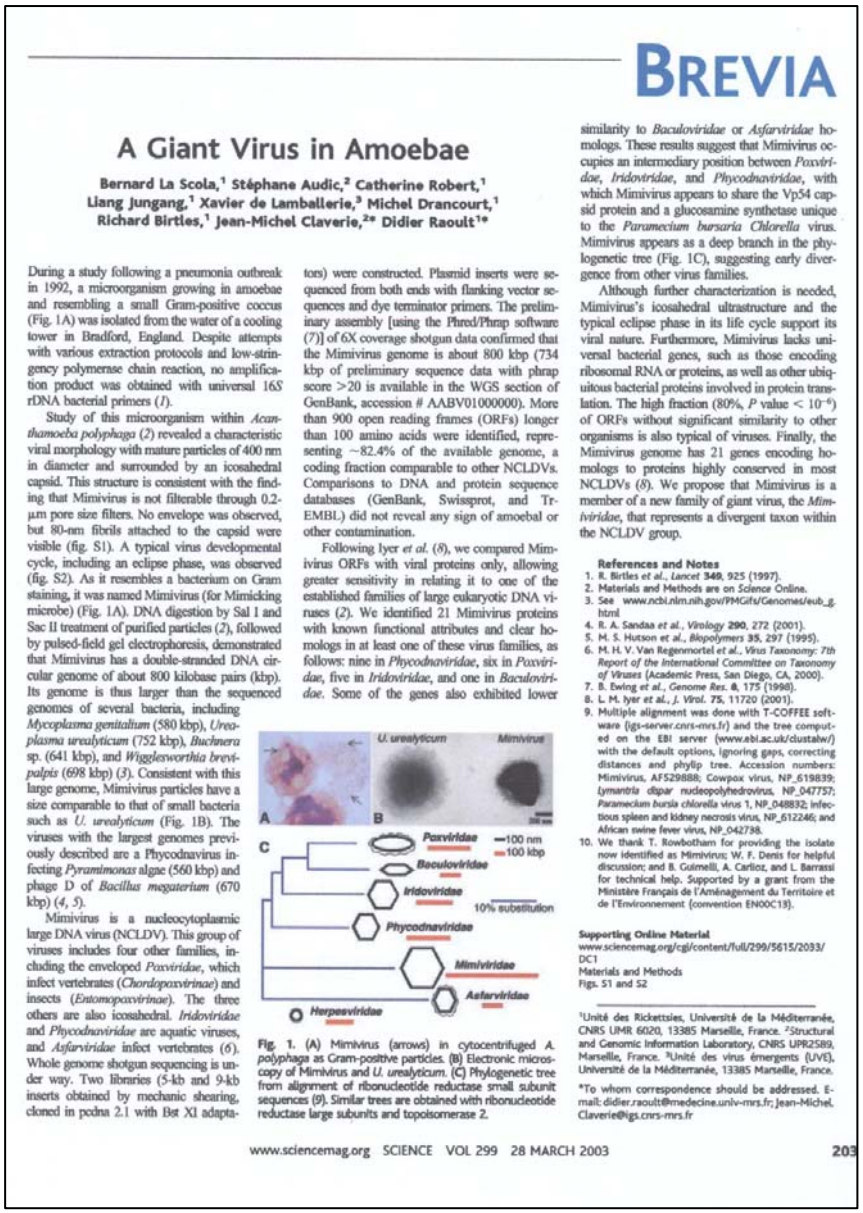

Bacillus subtilis Spore
Mimivirus
200 nm
0.2 µm

Viren sind normalerweise viel kleiner

- 20-300nm
- Mimivirus dagegen ist 600nm groß



(c) Garret Grisham



BREVIA

A Giant Virus in Amoebae

Bernard La Scola,[1] Stéphane Audic,[2] Catherine Robert,[1] Liang Jungang,[1] Xavier de Lamballerie,[3] Michel Drancourt,[1] Richard Birtles,[1] Jean-Michel Claverie,[2]* Didier Raoult[1]*

During a study following a pneumonia outbreak in 1992, a microorganism growing in amoebae and resembling a small Gram-positive coccus (Fig. 1A) was isolated from the water of a cooling tower in Bradford, England. Despite attempts with various extraction protocols and low-stringency polymerase chain reaction, no amplification product was obtained with universal 16*S* rDNA bacterial primers (*1*).

Study of this microorganism within *Acanthamoeba polyphaga* (*2*) revealed a characteristic viral morphology with mature particles of 400 nm in diameter and surrounded by an icosahedral capsid. This structure is consistent with the finding that Mimivirus is not filterable through 0.2-μm pore size filters. No envelope was observed, but 80-nm fibrils attached to the capsid were visible (fig. S1). A typical virus developmental cycle, including an eclipse phase, was observed (fig. S2). As it resembles a bacterium on Gram staining, it was named Mimivirus (for Mimicking microbe) (Fig. 1A). DNA digestion by Sal I and Sac II treatment of purified particles (*2*), followed by pulsed-field gel electrophoresis, demonstrated that Mimivirus has a double-stranded DNA circular genome of about 800 kilobase pairs (kbp). Its genome is thus larger than the sequenced genomes of several bacteria, including *Mycoplasma genitalium* (580 kbp), *Ureaplasma urealyticum* (752 kbp), *Buchnera* sp. (641 kbp), and *Wigglesworthia brevipalpis* (698 kbp) (*3*). Consistent with this large genome, Mimivirus particles have a size comparable to that of small bacteria such as *U. urealyticum* (Fig. 1B). The viruses with the largest genomes previously described are a Phycodnavirus infecting *Pyramimonas* algae (560 kbp) and phage D of *Bacillus megaterium* (670 kbp) (*4*, *5*).

Mimivirus is a nucleocytoplasmic large DNA virus (NCLDV). This group of viruses includes four other families, including the enveloped *Poxviridae*, which infect vertebrates (*Chordopoxvirinae*) and insects (*Entomopoxvirinae*). The three others are also icosahedral. *Iridoviridae* and *Phycodnaviridae* are aquatic viruses, and *Asfarviridae* infect vertebrates (*6*). Whole genome shotgun sequencing is under way. Two libraries (5-kb and 9-kb inserts obtained by mechanic shearing, cloned in pedna 2.1 with Bst XI adaptators) were constructed. Plasmid inserts were sequenced from both ends with flanking vector sequences and dye terminator primers. The preliminary assembly [using the Phred/Phrap software (*7*)] of 6X coverage shotgun data confirmed that the Mimivirus genome is about 800 kbp (734 kbp of preliminary sequence data with phrap score >20 is available in the WGS section of GenBank, accession # AABV01000000). More than 900 open reading frames (ORFs) longer than 100 amino acids were identified, representing ~82.4% of the available genome, a coding fraction comparable to other NCLDVs. Comparisons to DNA and protein sequence databases (GenBank, Swissprot, and TrEMBL) did not reveal any sign of amoebal or other contamination.

Following Iyer *et al.* (*8*), we compared Mimivirus ORFs with viral proteins only, allowing greater sensitivity in relating it to one of the established families of large eukaryotic DNA viruses (*2*). We identified 21 Mimivirus proteins with known functional attributes and clear homologs in at least one of these virus families, as follows: nine in *Phycodnaviridae*, six in *Poxviridae*, five in *Iridoviridae*, and one in *Baculoviridae*. Some of the genes also exhibited lower similarity to *Baculoviridae* or *Asfarviridae* homologs. These results suggest that Mimivirus occupies an intermediary position between *Poxviridae*, *Iridoviridae*, and *Phycodnaviridae*, with which Mimivirus appears to share the Vp54 capsid protein and a glucosamine synthetase unique to the *Paramecium bursaria Chlorella* virus. Mimivirus appears as a deep branch in the phylogenetic tree (Fig. 1C), suggesting early divergence from other virus families.

Although further characterization is needed, Mimivirus's icosahedral ultrastructure and the typical eclipse phase in its life cycle support its viral nature. Furthermore, Mimivirus lacks universal bacterial genes, such as those encoding ribosomal RNA or proteins, as well as other ubiquitous bacterial proteins involved in protein translation. The high fraction (80%, P value $< 10^{-6}$) of ORFs without significant similarity to other organisms is also typical of viruses. Finally, the Mimivirus genome has 21 genes encoding homologs to proteins highly conserved in most NCLDVs (*8*). We propose that Mimivirus is a member of a new family of giant virus, the *Mimiviridae*, that represents a divergent taxon within the NCLDV group.

Fig. 1. (A) Mimivirus (arrows) in cytocentrifuged *A. polyphaga* as Gram-positive particles. (B) Electronic microscopy of Mimivirus and *U. urealyticum*. (C) Phylogenetic tree from alignment of ribonucleotide reductase small subunit sequences (*9*). Similar trees are obtained with ribonucleotide reductase large subunits and topoisomerase 2.

References and Notes

1. R. Birtles *et al.*, *Lancet* **340**, 925 (1997).
2. Materials and Methods are on *Science* Online.
3. See www.ncbi.nlm.nih.gov/PMGifs/Genomes/eub_g.html
4. R. A. Sandaa *et al.*, *Virology* **290**, 272 (2001).
5. M. S. Hutson *et al.*, *Biopolymers* **35**, 297 (1995).
6. M. H. V. Van Regenmortel *et al.*, *Virus Taxonomy: 7th Report of the International Committee on Taxonomy of Viruses* (Academic Press, San Diego, CA, 2000).
7. B. Ewing *et al.*, *Genome Res.* **8**, 175 (1998).
8. L. M. Iyer *et al.*, *J. Virol.* **75**, 11720 (2001).
9. Multiple alignment was done with T-COFFEE software (igs-server.cnrs-mrs.fr) and the tree computed on the EBI server (www.ebi.ac.uk/clustalw/) with the default options, ignoring gaps, correcting distances and phylip tree. Accession numbers: Mimivirus, AF529888; Cowpox virus, NP_619839; *Lymantria dispar* nucleopolyhedrovirus, NP_047757; *Paramecium bursla chlorella* virus 1, NP_048832; infectious spleen and kidney necrosis virus, NP_612246; and African swine fever virus, NP_042738.
10. We thank T. Rowbotham for providing the isolate now identified as Mimivirus; W. F. Denis for helpful discussion; and B. Guimelli, A. Carlioz, and L. Barrassi for technical help. Supported by a grant from the Ministère Français de l'Aménagement du Territoire et de l'Environnement (convention EN00C13).

Supporting Online Material

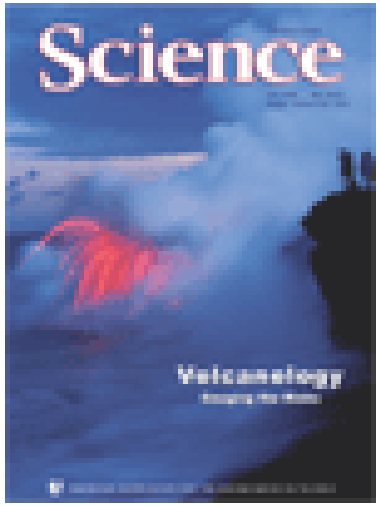
www.sciencemag.org/cgi/content/full/299/5615/2033/DC1
Materials and Methods
Figs. S1 and S2

[1]Unité des Rickettsies, Université de la Méditerranée, CNRS UMR 6020, 13385 Marseille, France. [2]Structural and Genomic Information Laboratory, CNRS UPR2589, Marseille, France. [3]Unité des virus émergents (UVE), Université de la Méditerranée, 13385 Marseille, France.

*To whom correspondence should be addressed. E-mail: didier.raoult@medecine.univ-mrs.fr; Jean-Michel.Claverie@igs.cnrs-mrs.fr

1. Veröffentlichung:

La Scola et al.
Science **299**, p. 233
(March 2003)



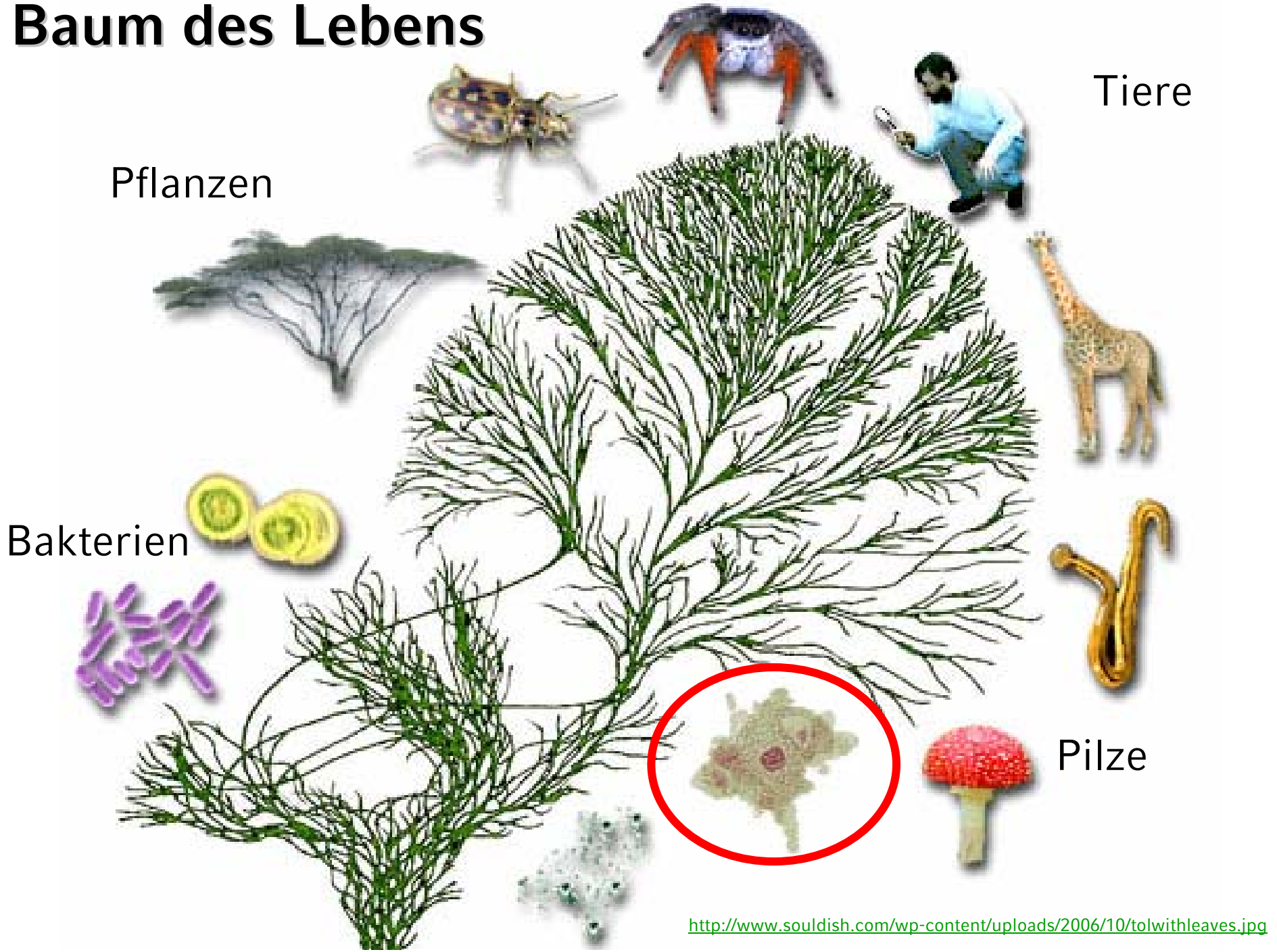
Die zentralen Fragestellungen :

Was für ein „Wesen“ ist Mimivirus?

Wo ist der Platz von Mimivirus in der Evolution?

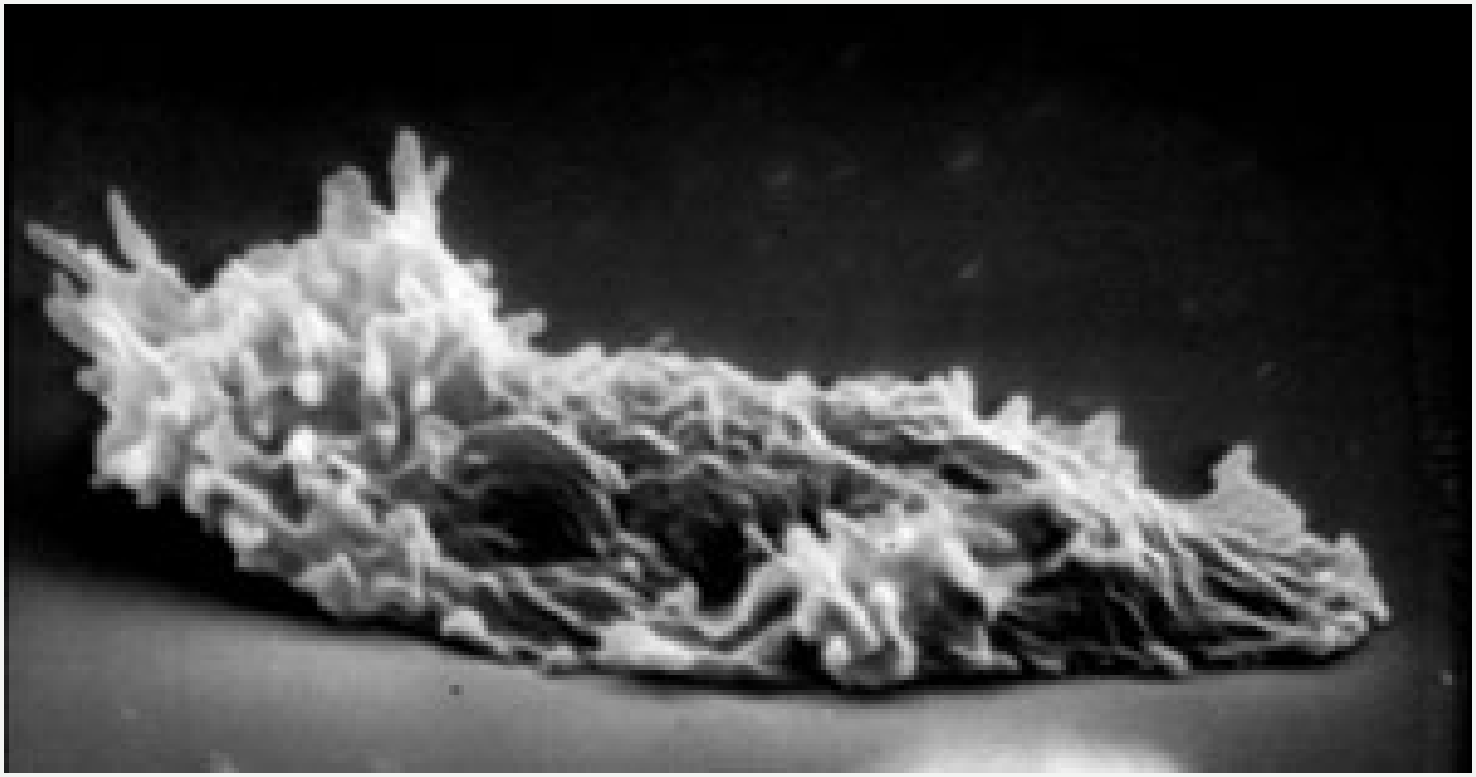
Warum ist es so groß?

Baum des Lebens



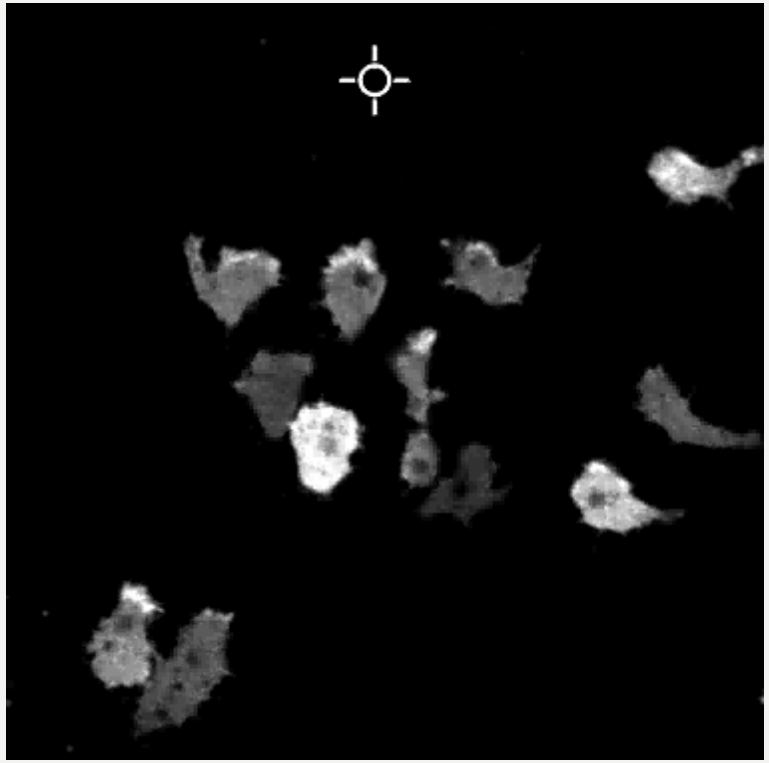
http://www.souldish.com/wp-content/uploads/2006/10/tolwithleaves.jpg

Die Amöbe *Acanthamoeba castellanii*



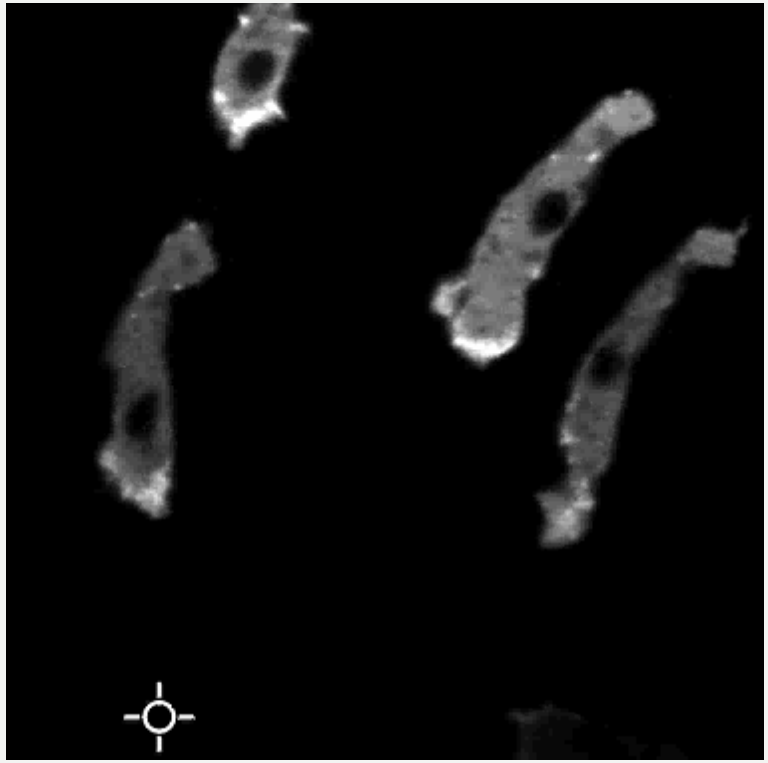
***Acanthamoeba castellanii* (c) Steve Doberstein**

Die Amöbe *Dictyostelium discoideum*

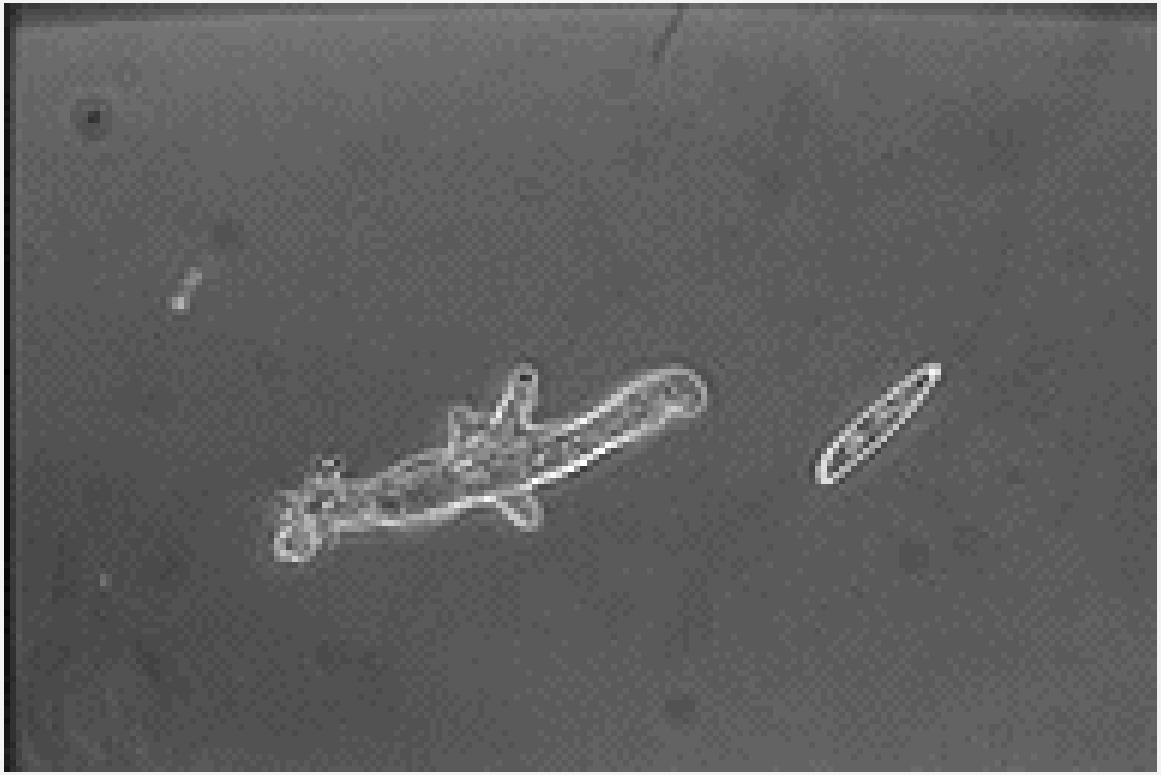


K. Barisic, M. Ecke, C. Heizer, M. Maniak,
M. Westphal, R. Albrecht, G. Gerisch,
Max-Planck-Institut fur Biochemie, Martinsried.

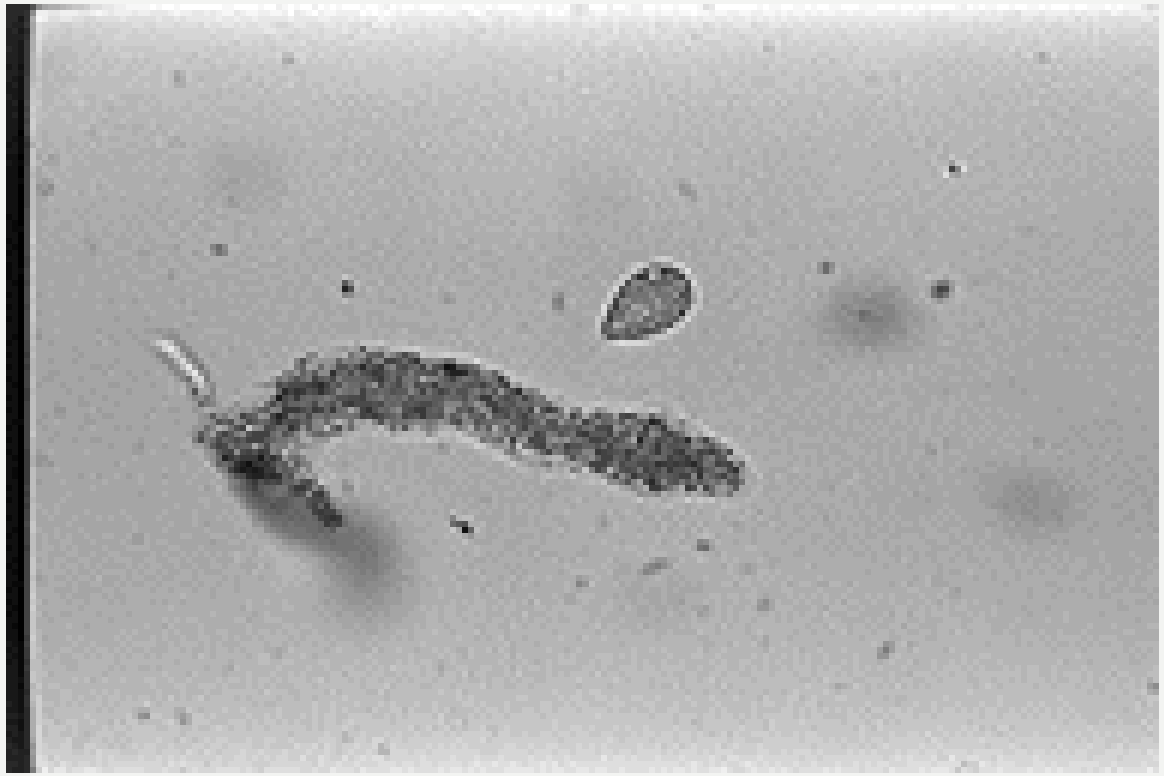
Die Amöbe *Dictyostelium discoideum*



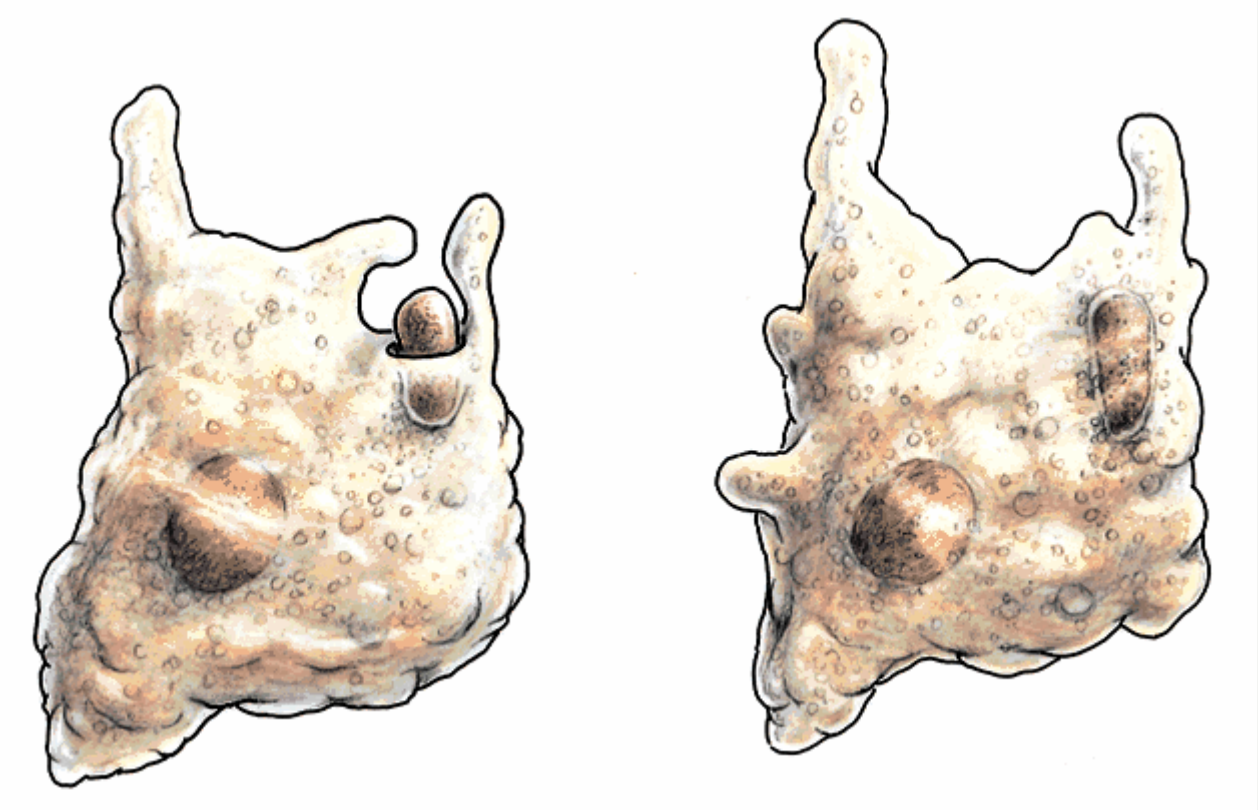
K. Barisic, M. Ecke, C. Heizer, M. Maniak,
M. Westphal, R. Albrecht, G. Gerisch,
Max-Planck-Institut fur Biochemie, Martinsried.



http://academic.brooklyn.cuny.edu/biology/bio4fv/page/phago.htm



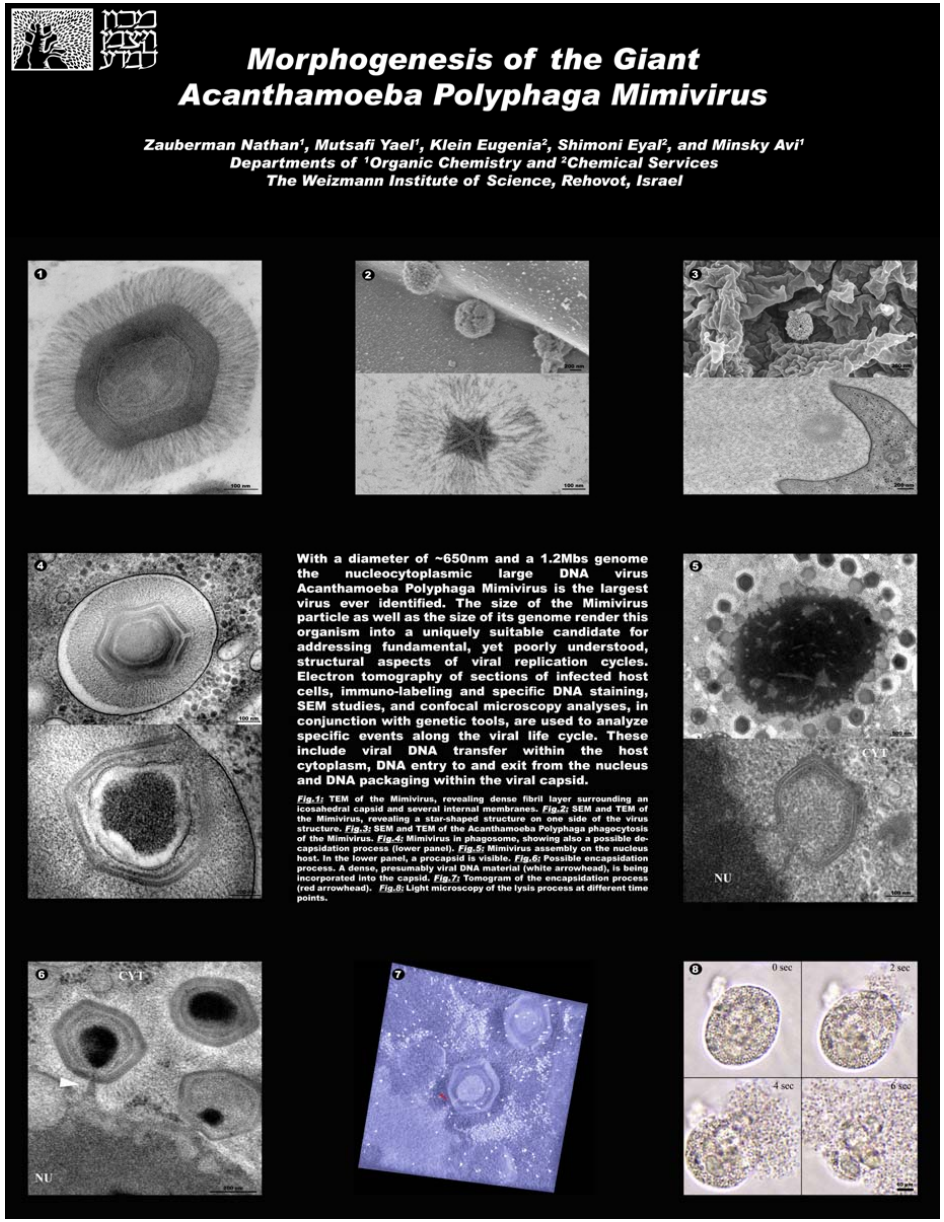
http://academic.brooklyn.cuny.edu/biology/bio4fv/page/phago.htm



http://www.agen.ufl.edu/~chyn/age2062/lect/lect_06/4_25B.GIF

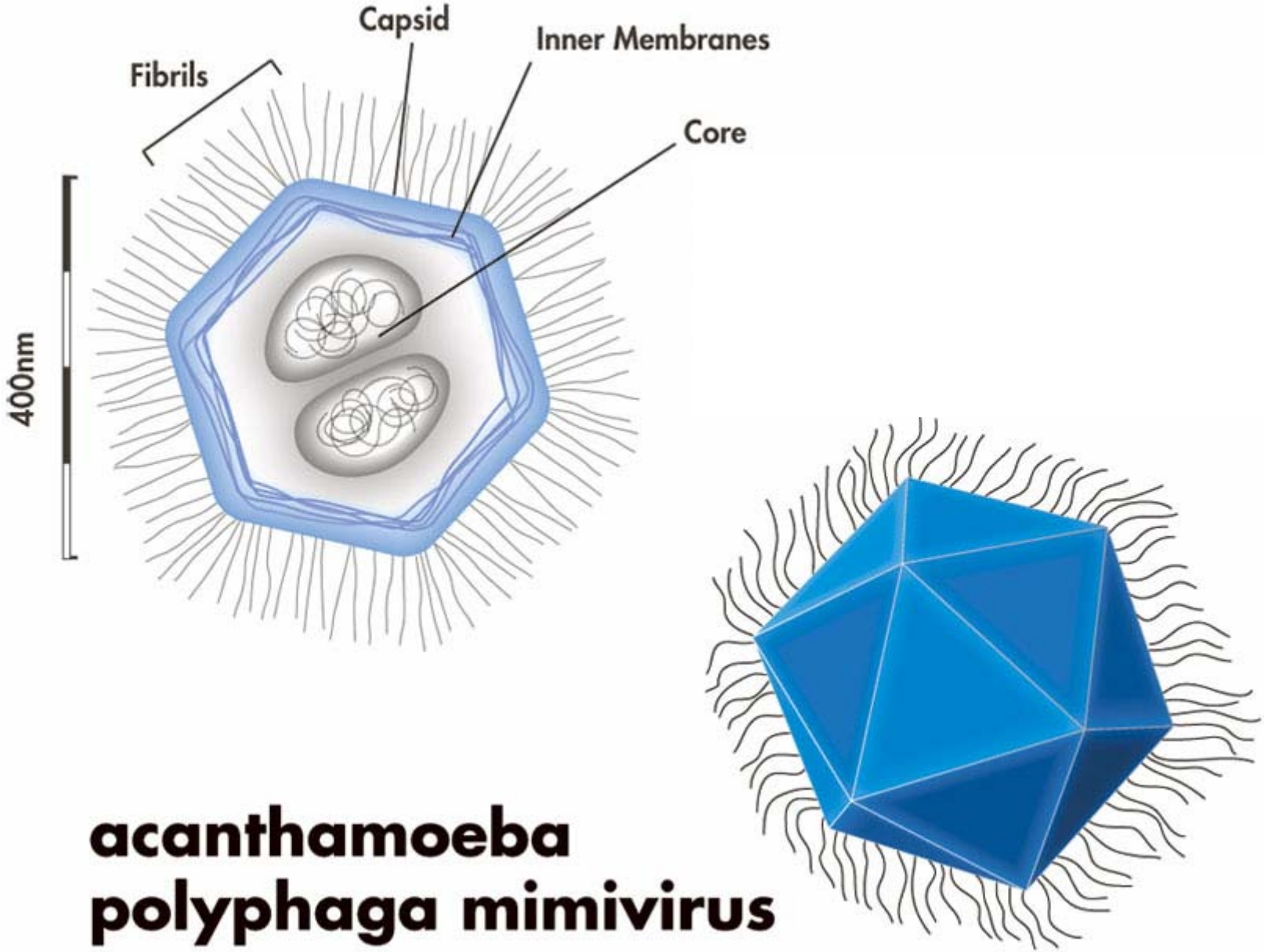
„Phagozytose“

- Amöben ernähren sich von Bakterien
- Amöben sind Räuber
- Amöben schließen Bakterien ein und verdauen sie
- ... aber nicht alle Nahrung ist gut verdaulich!

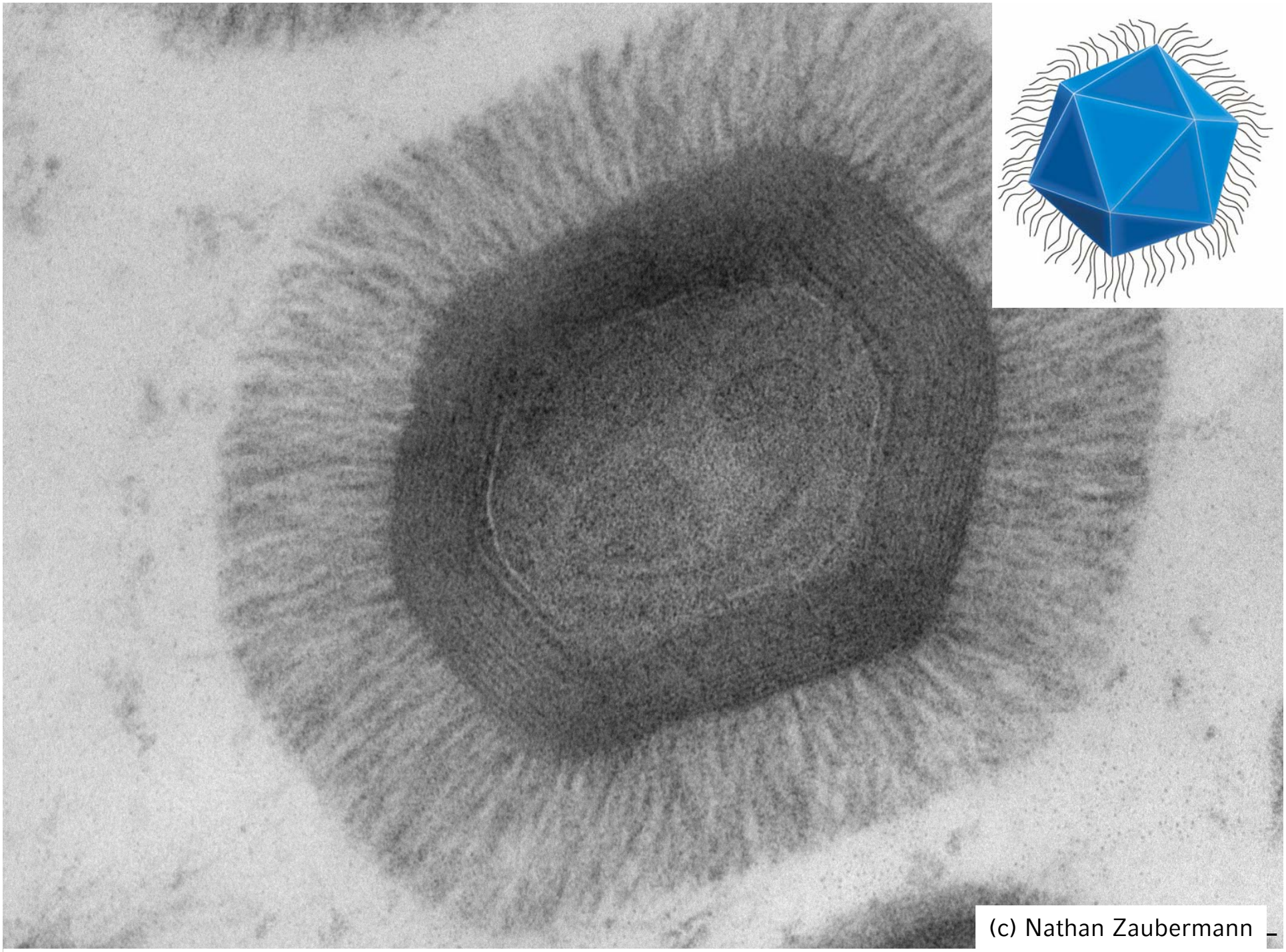


Nathan Zaubermann
Weizmann Institut
Israel

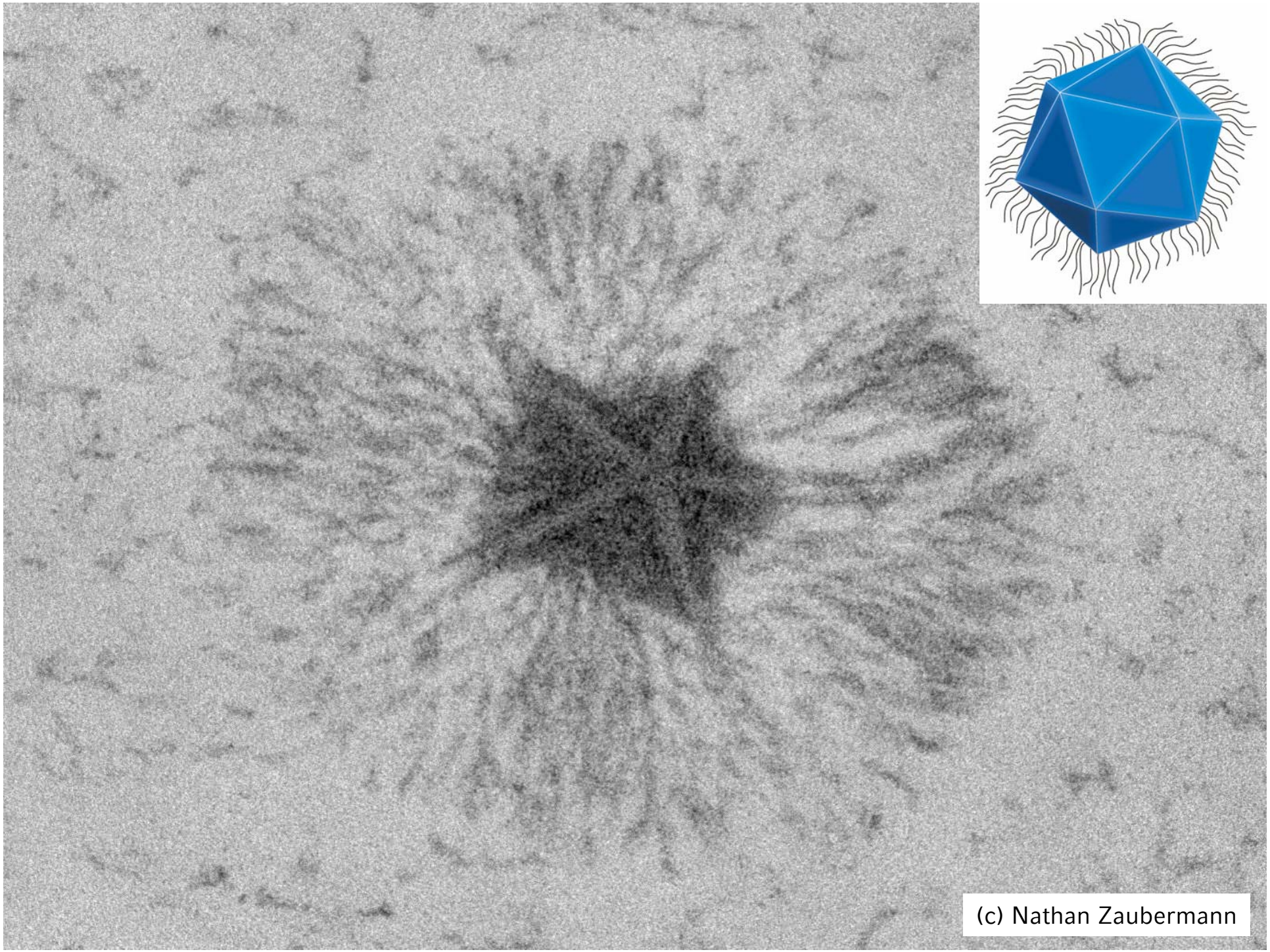
http://www.weizmann.ac.il/Organic_Chemistry/minsky/nathan/nathan.shtml

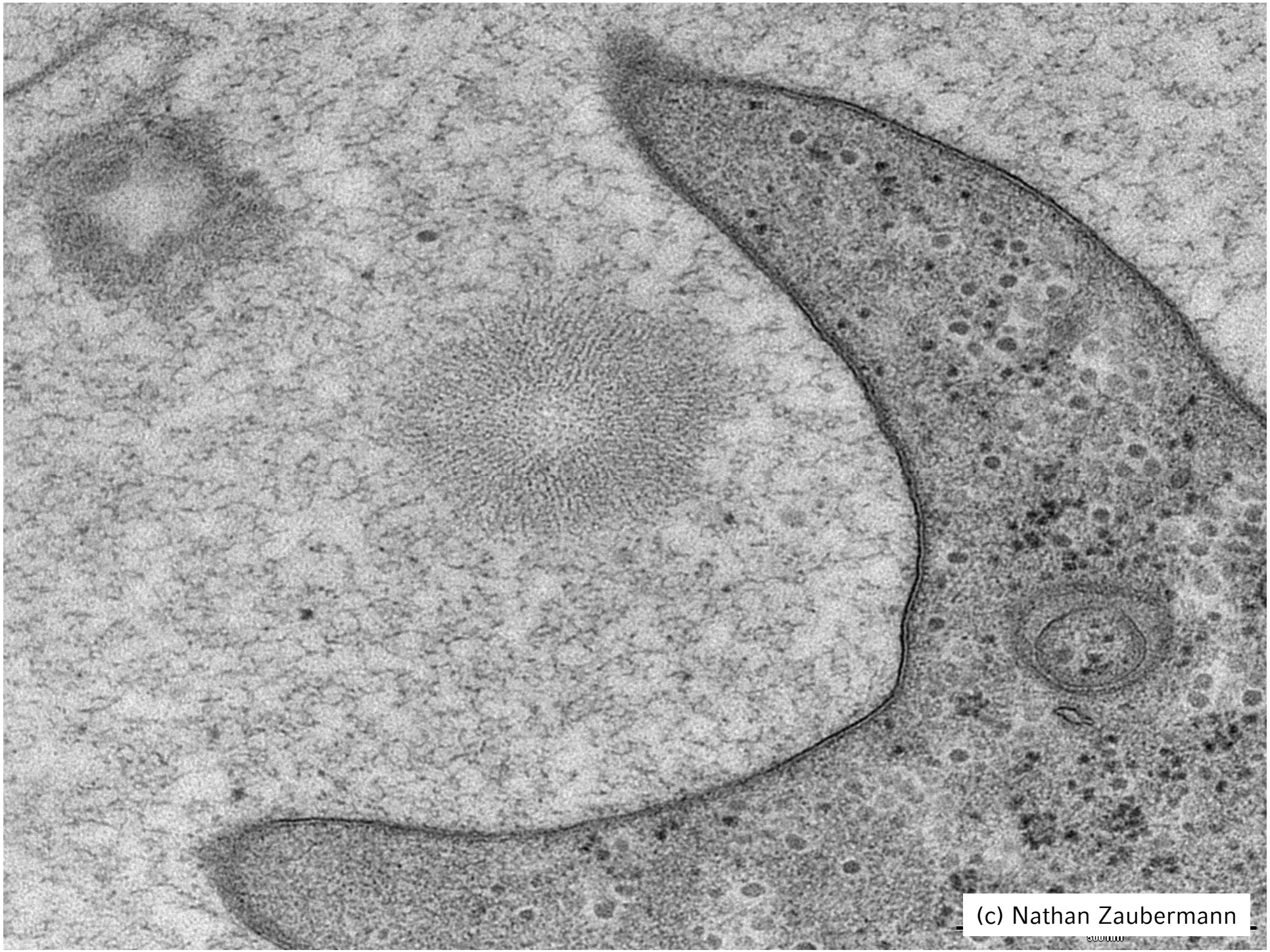

Capsid
Inner Membranes
Fibrils
Core
400nm
acanthamoeba
polyphaga mimivirus

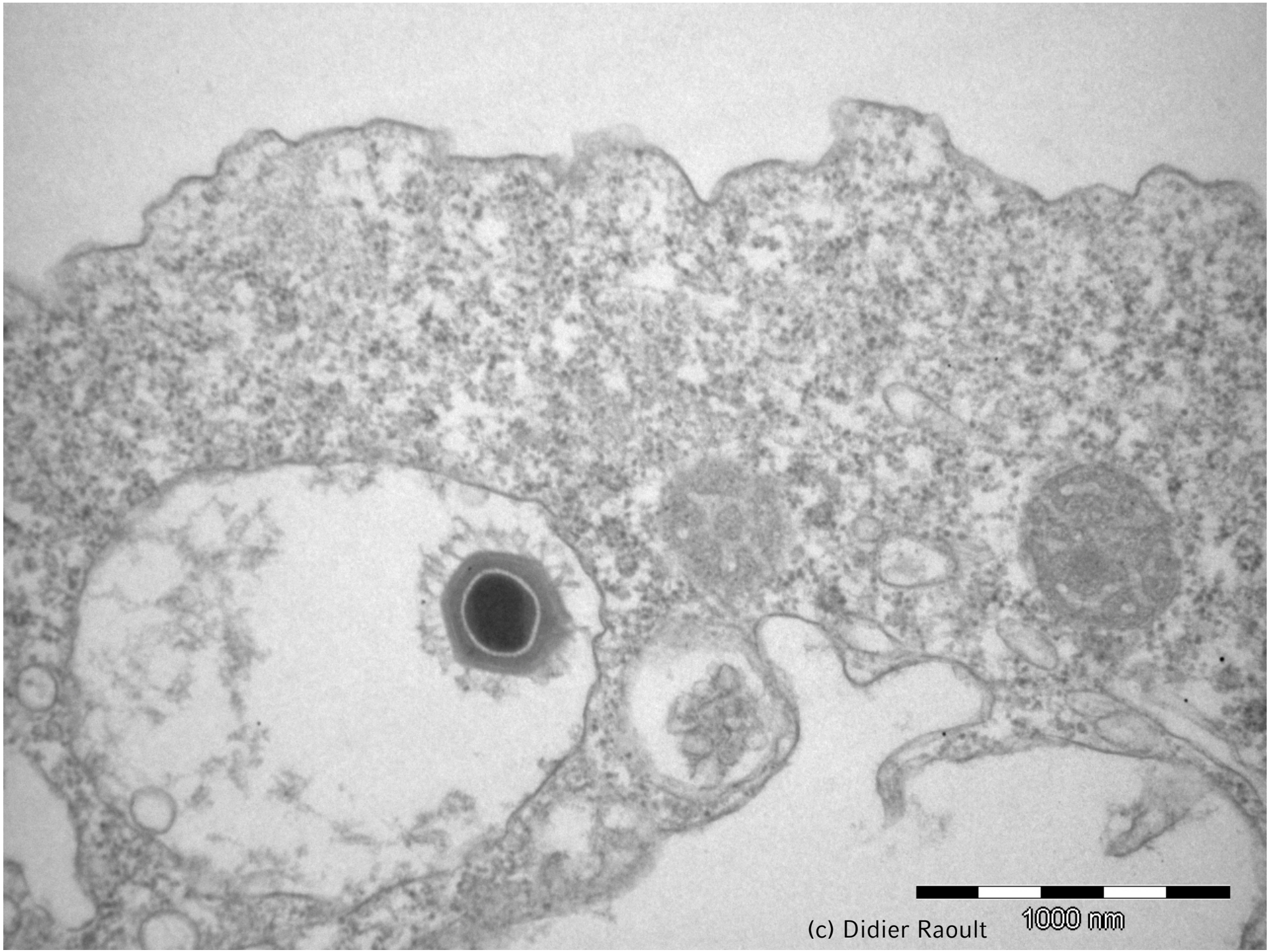
(c) Wikipedia

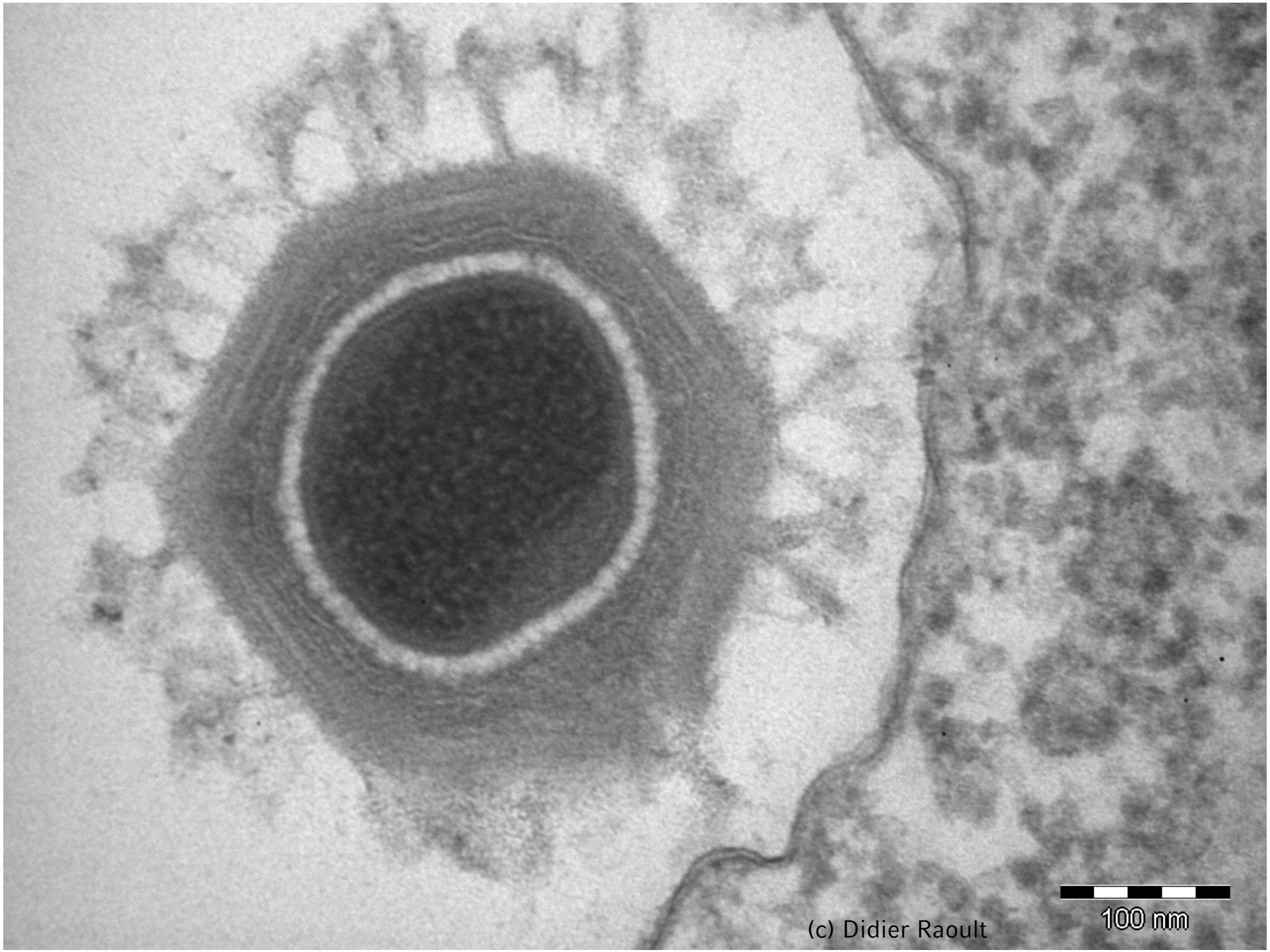


(c) Nathan Zaubermann

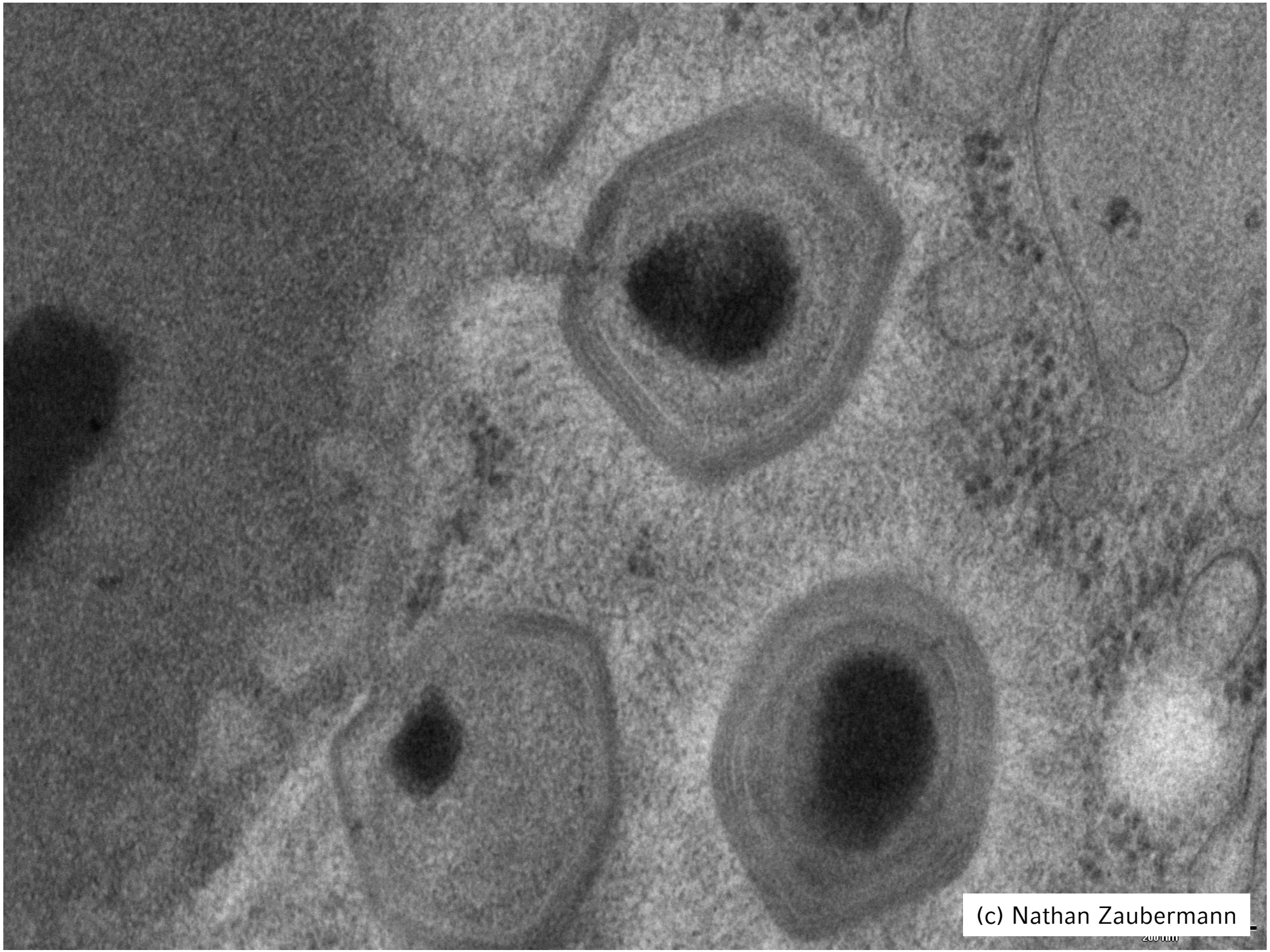

(c) Nathan Zaubermann

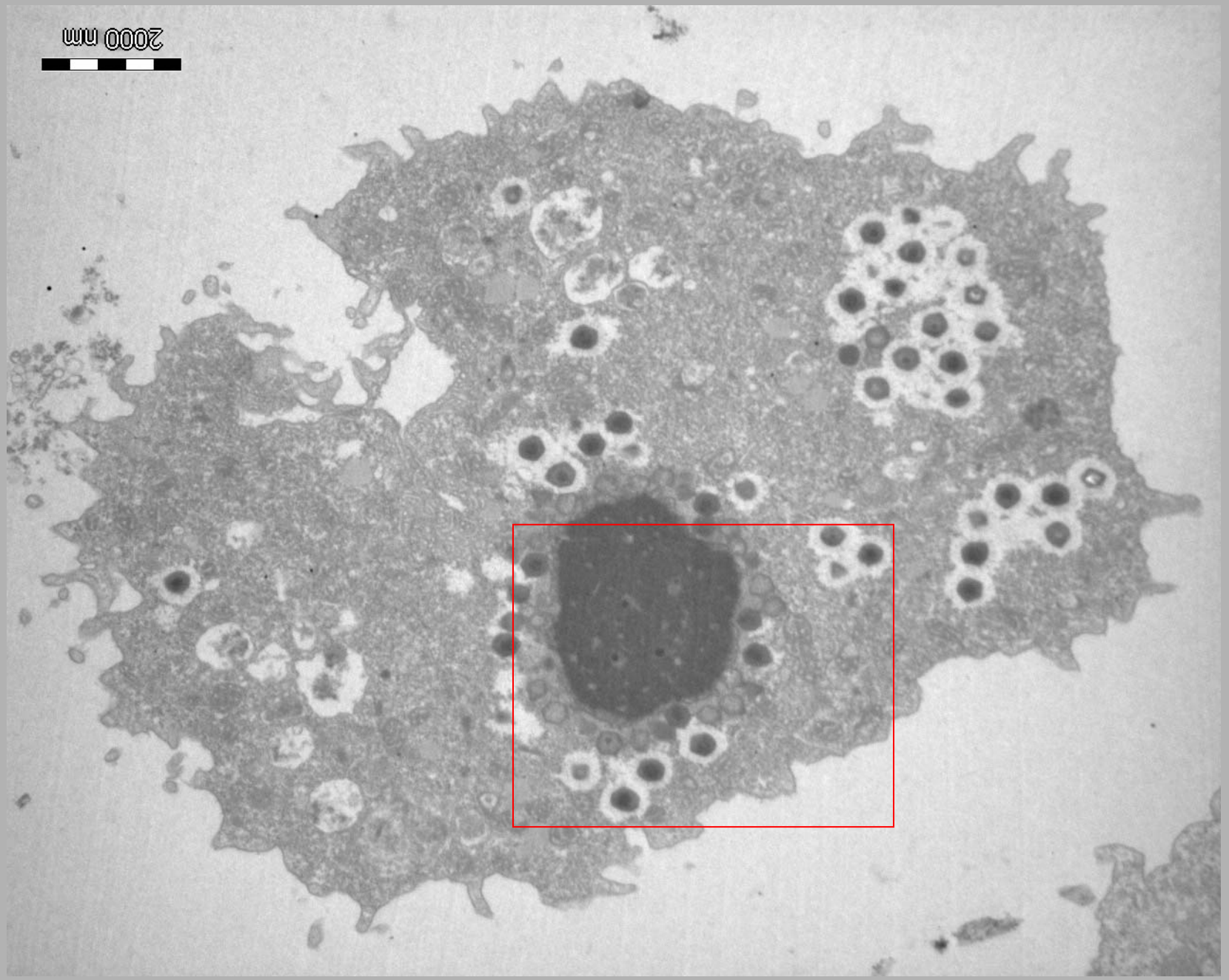

(c) Nathan Zaubermann

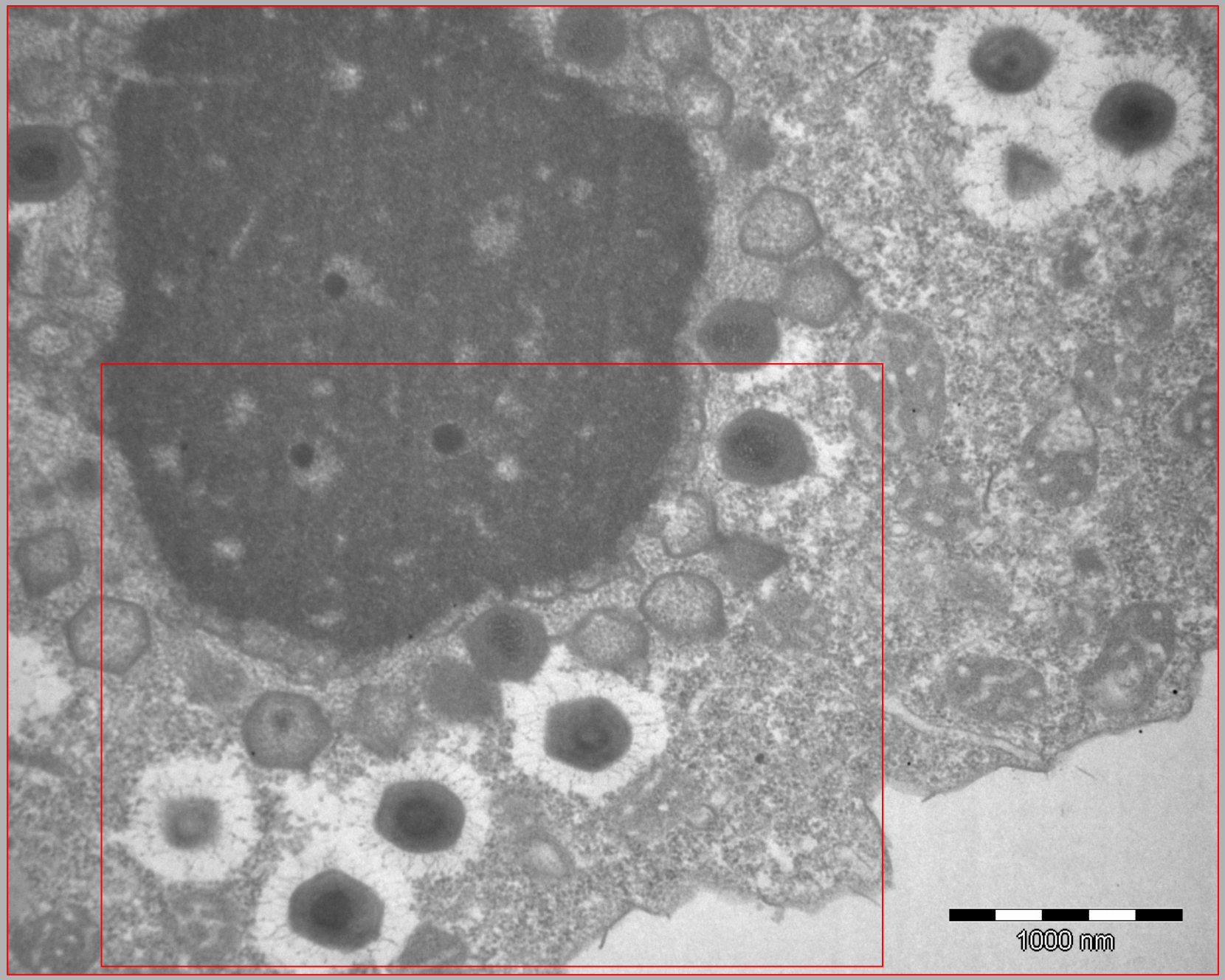

1000 nm
(c) Didier Raoult

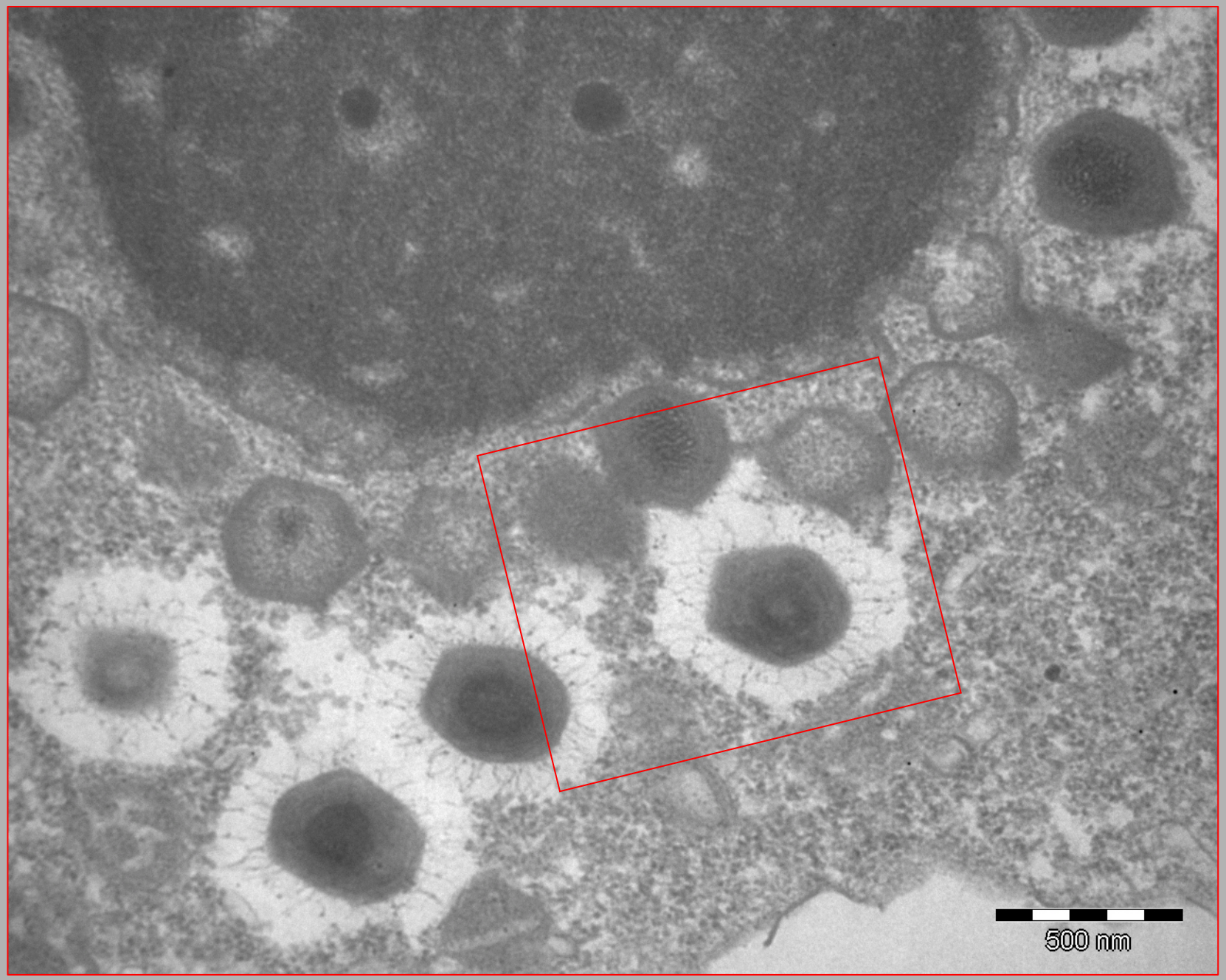

100 nm
(c) Didier Raoult

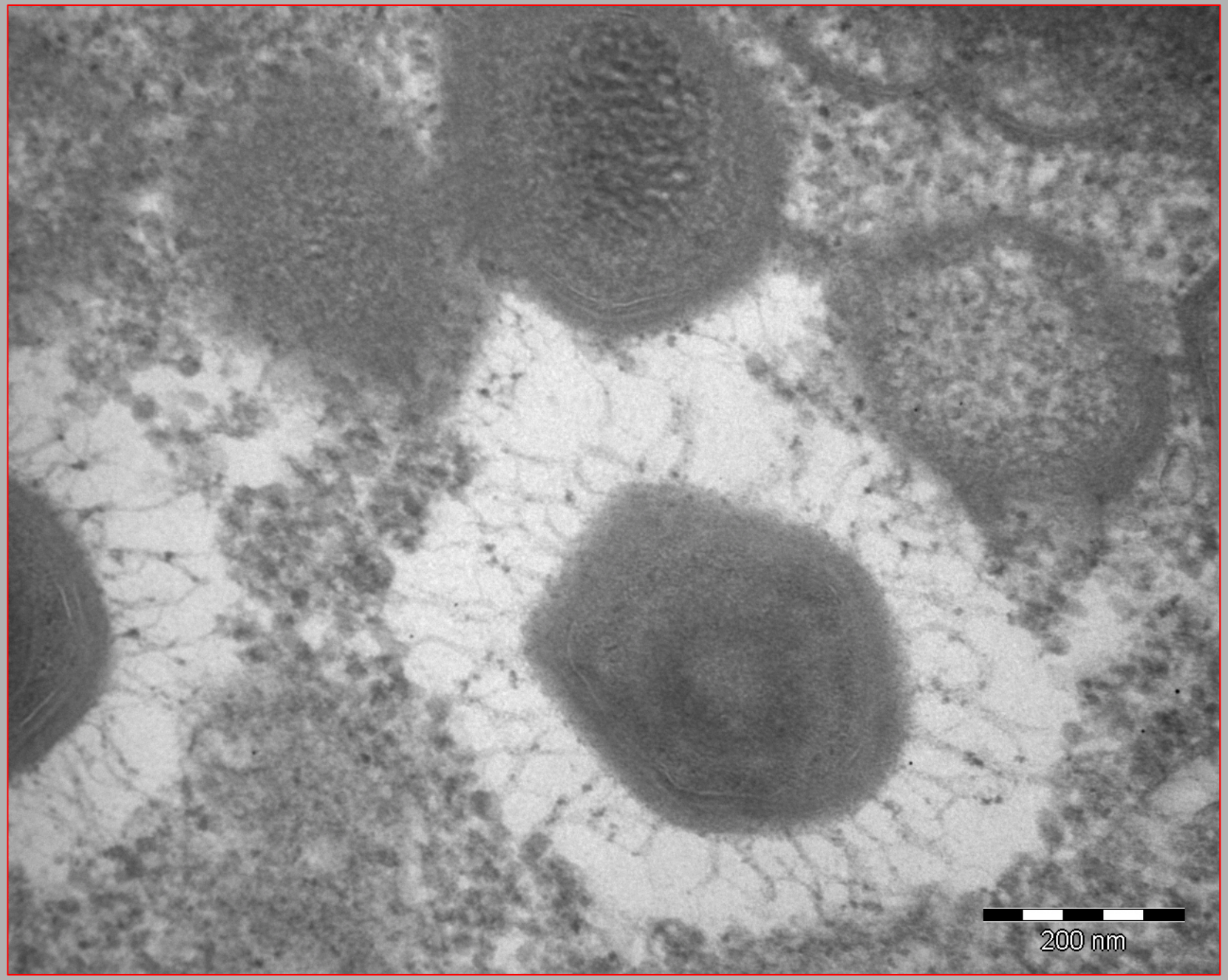
ECLIPSE PHASE


(c) Nathan Zaubermann









Wie kann man mehr über Mimivirus erfahren?

Genomsequenzierung ...

... und bioinformatische Analyse des Genoms

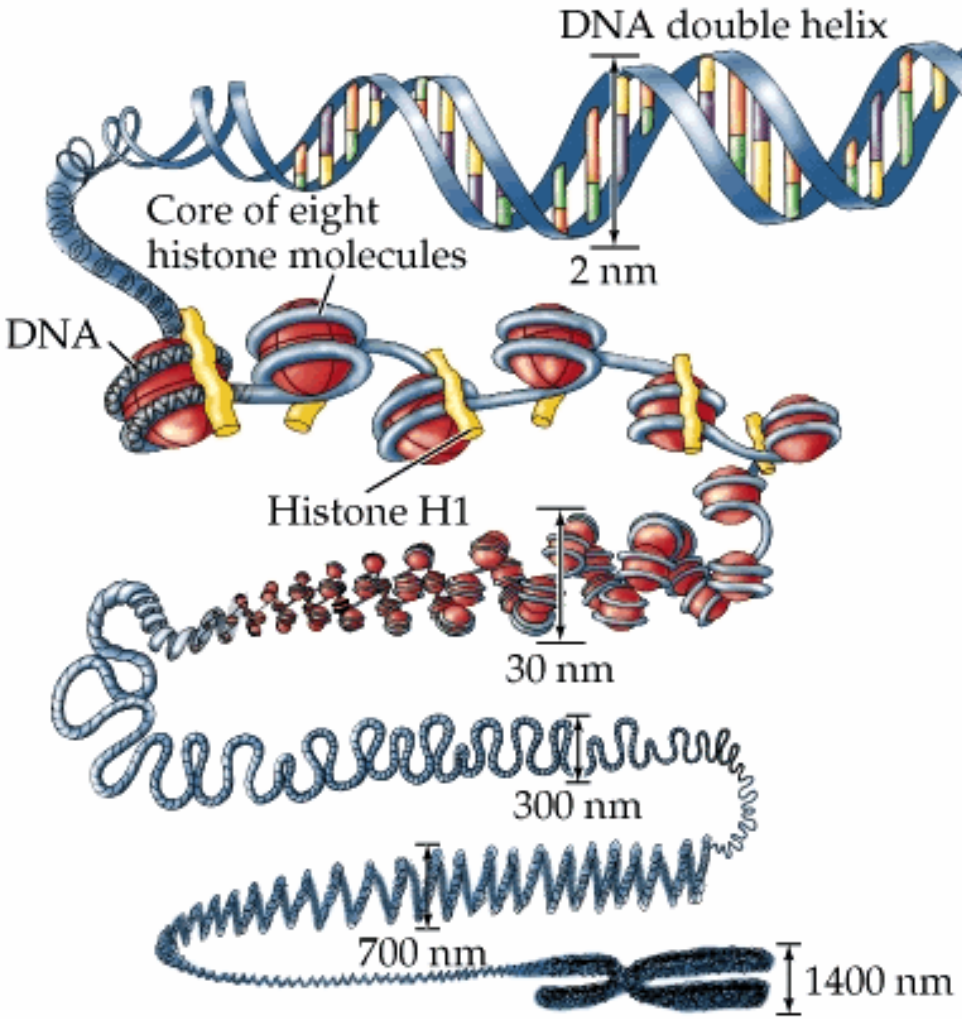
Bioinformatik ist die Wissenschaft, die sich (unter anderem) mit der Analyse von Genomsequenzen befasst.

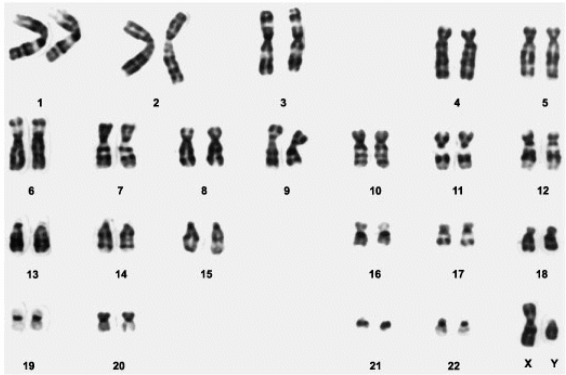
Alle Lebewesen haben eines gemeinsam:

ihr Erbgut ist in der DNA verschlüsselt

Francis Crick and James Watson point out features of their model for the structure of DNA. (© A. Barrington Brown/Science Source/Photo Researchers, Inc.)

(c) Garret Grisham

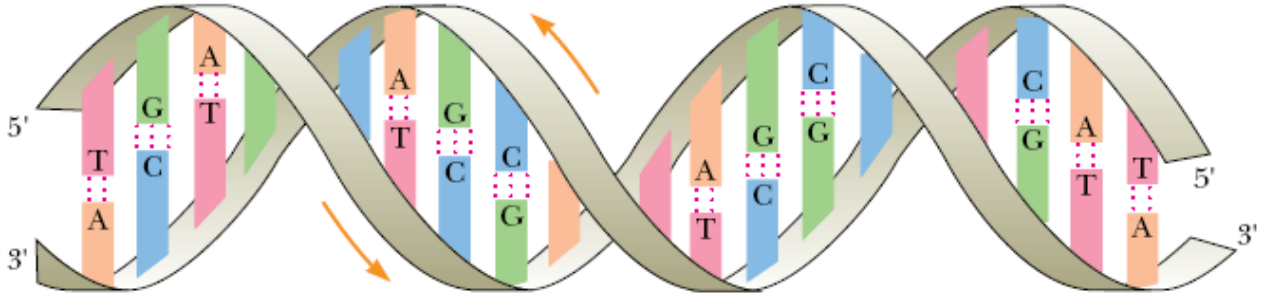


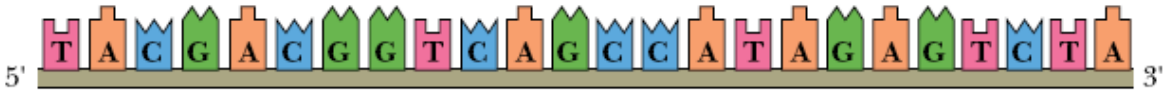


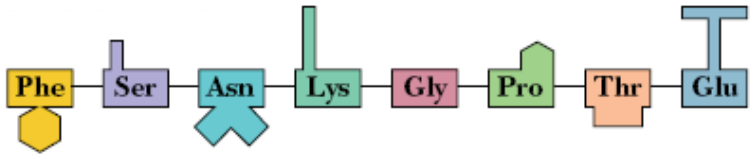
Das Erbgut von Mimivirus wurde entschlüsselt

- Man fand fast 1.2 Millionen Basenpaare (Nukleotide).
- Das ist mehr als doppelt so groß wie das Genom der kleinsten frei lebenden Bakterien.
- 911 Gene wurden identifiziert.

Das zentrale Paradigma der Biologie







DNA

RNA

Protein

(c) Garret Grisham

Es gibt zwei „Alphabete“ in der Biologie

4 Buchstaben für DNA und RNA (4 Nukleotide)
A-T-G-C (RNA: A-U-G-C)

20 Buchstaben für Proteine (20 Aminosäuren)
ACDEFGHIKLMNPQRSTVWY

Ein Mimivirus Gen in FASTA Format (DNA Kode)

```
>MIMI_L176  genome:mim.genome pos:c 212394..213530
atgggtgattttgtattttgttatggttctcataatagaaaactcttttc
aaaatatattgtgagtacaaaaaatagtcacaaattttatttcggcaaaa
aacaatatatcgatattgaaatagaaacaatttttagtcaaaaatacatt
gtggtagatttcaatgacagtctcgatttcttgaattttattgttgacaa
tgaaatttattgtaacttcacaaacaacaaacacgaatgtgaatcattag
aatttcaccataattacttgaattatataacgcaacacaaacacttggat
[...]
tatcaaggaattaatcaaagaaactagtgactcataa
```

Ein Mimivirus Gen in FASTA Format (übersetzt in Protein Kode)

```
>MIMI_L176 complement(212394..213530) Unknown
MGDFVFCYGSHNRKLFSKYIVSTKNSHKFYFGKKQYIDIEIETIFSQKYI
VVDFNDSLDFLNFIVDNEIYCNFTNNKHECESLEFHHNYLNYITQHKHLD
IVKIFYKKFVPLIKSGRGVQSLQFCILQDIDPEVVKCVFKNGSLEDTKDF
IINEFHKNPDITIEFMDEIISIYKHKLTKLFMENKIDKSIDNMEISLFQF
LIPALKKDDVGLFNFIIEEICNLTSEIDKTKLDKTQLGYLESIDIDFGIR
DINTLIDYYMLCDFDDCSPEDEMYFCPNIFRQLIFSLDNLDSLAGRILFD
IPEYNVVEYVGIICDFIGETNPLLINKMLPEAKSTEMAQLLIDCGADYEK
LYESKKFSKCNSCVKKLIKELIKETSDS
```

Proteine sind die Bausteine des Lebens

- Struktur (z.B. in Haar, Haut, Muskel)
- Enzyme (Verdauung, Aufbau wichtiger Moleküle)
- Signale (Hormone)
- Toxine (z.B. Schlangengift)
- Verteidigung (Antikörper)
- Kopie der DNA
- Regulatorische Aufgaben
- ... und vieles mehr

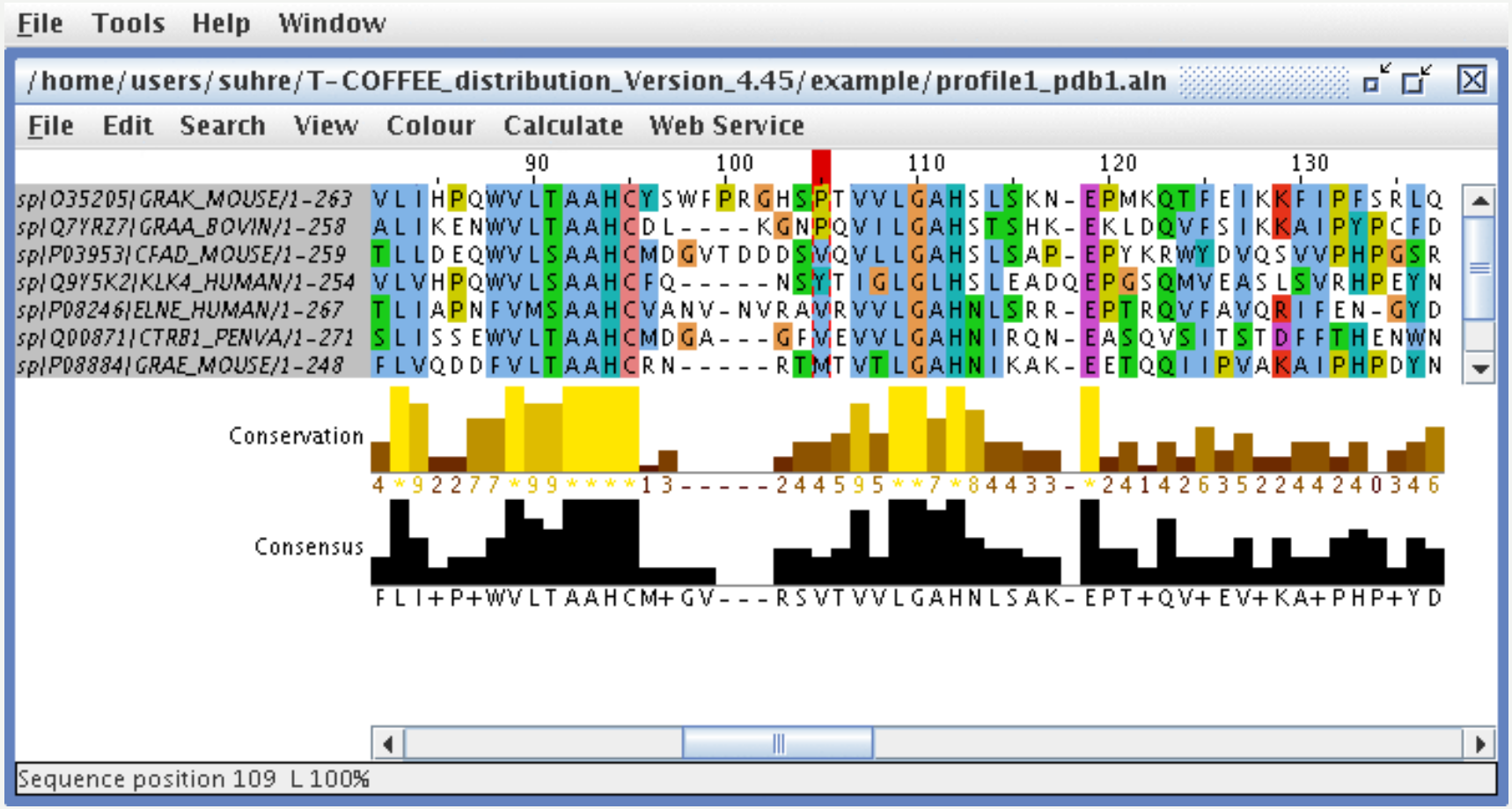
Die Bestimmung von Proteinfunktionen ist eine Hauptaufgabe der Biologie

- Molekularbiologie / Experimente
  ... an Modelorganismen wie *Escherichia coli*
- Man weiß ...
  ... ähnliche Proteine haben ähnliche Funktion
- „Homologie“

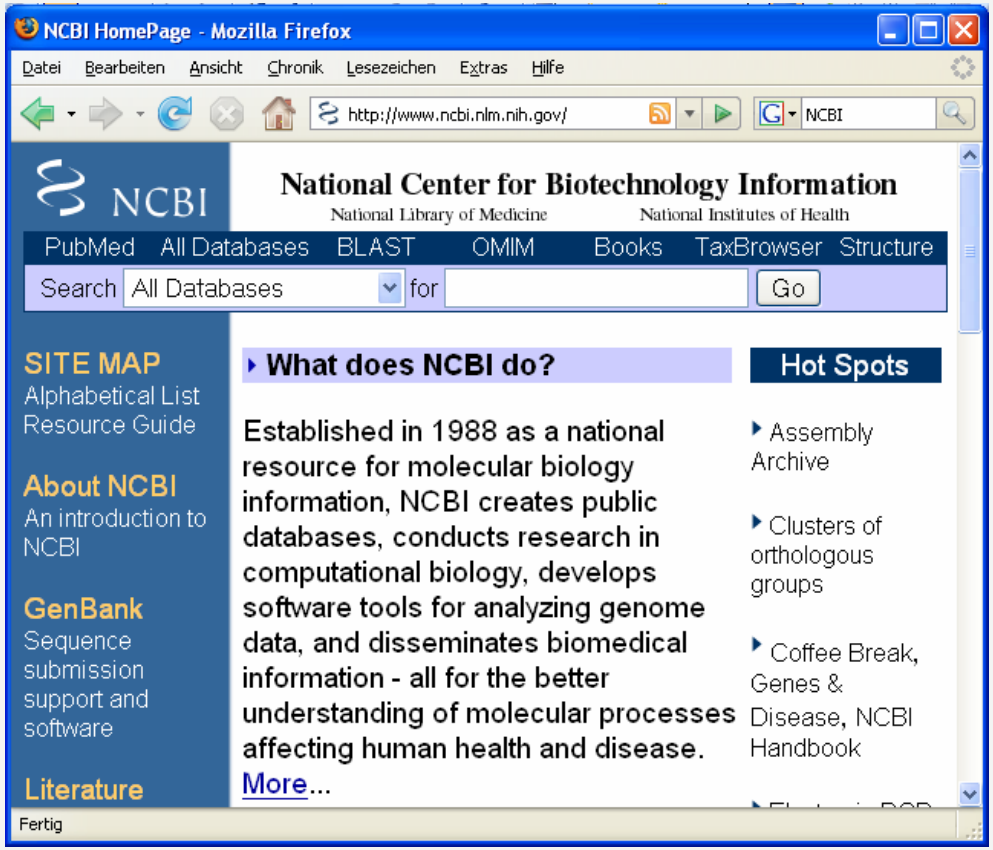
Mit Hilfe der Bioinformatik kann Wissen über bekannte Organismen auf Unbekannte übertragen werden

- Sequenzvergleich ist eine Hauptaufgabe der Bioinformatik

Ein multiples Alignment



Die NCBI Bioinformatik Datenbank

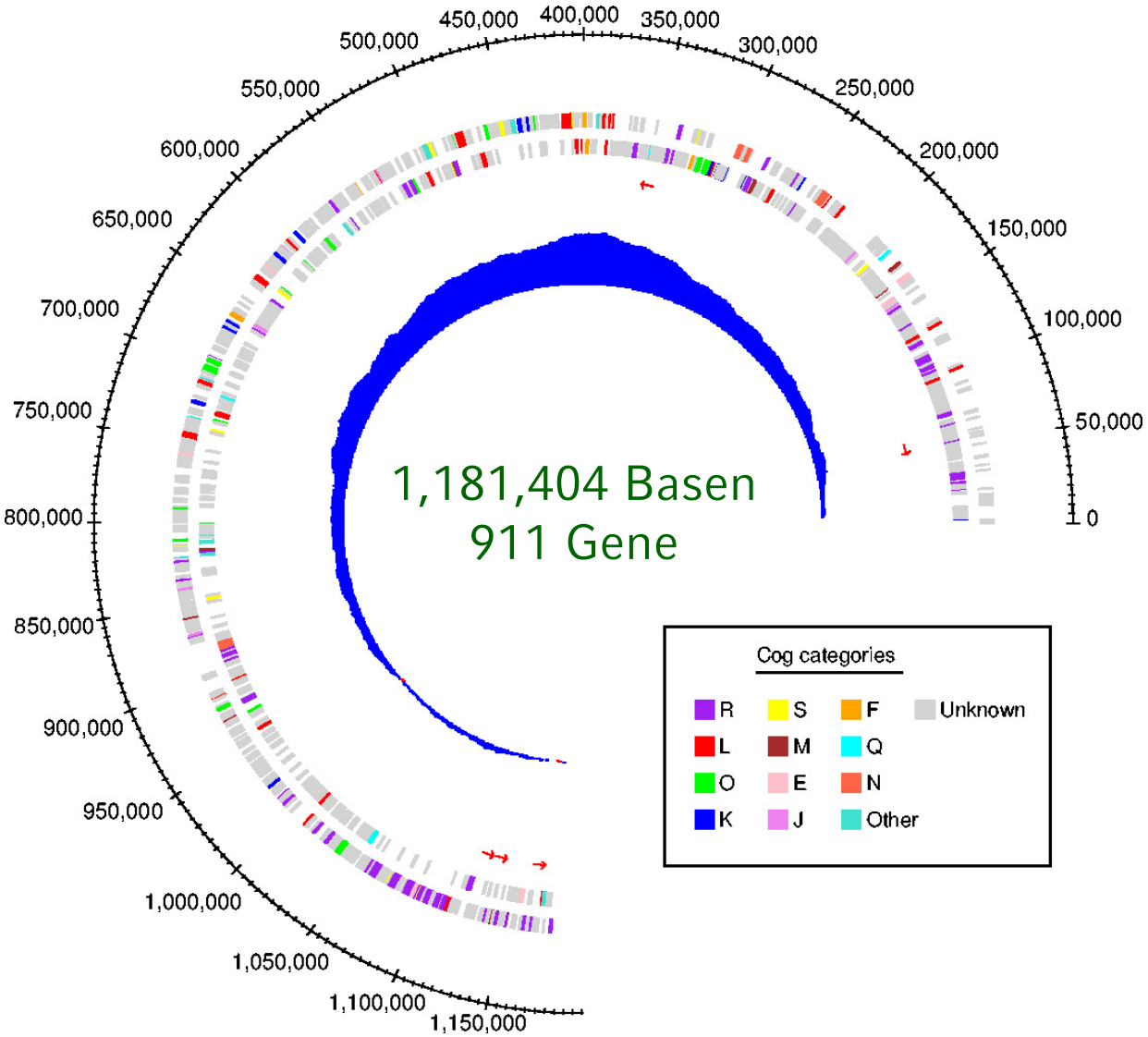


http://www.ncbi.nlm.nih.gov/

Somit wurden völlig neue Proteinfunktionen im Mimivirus entdeckt

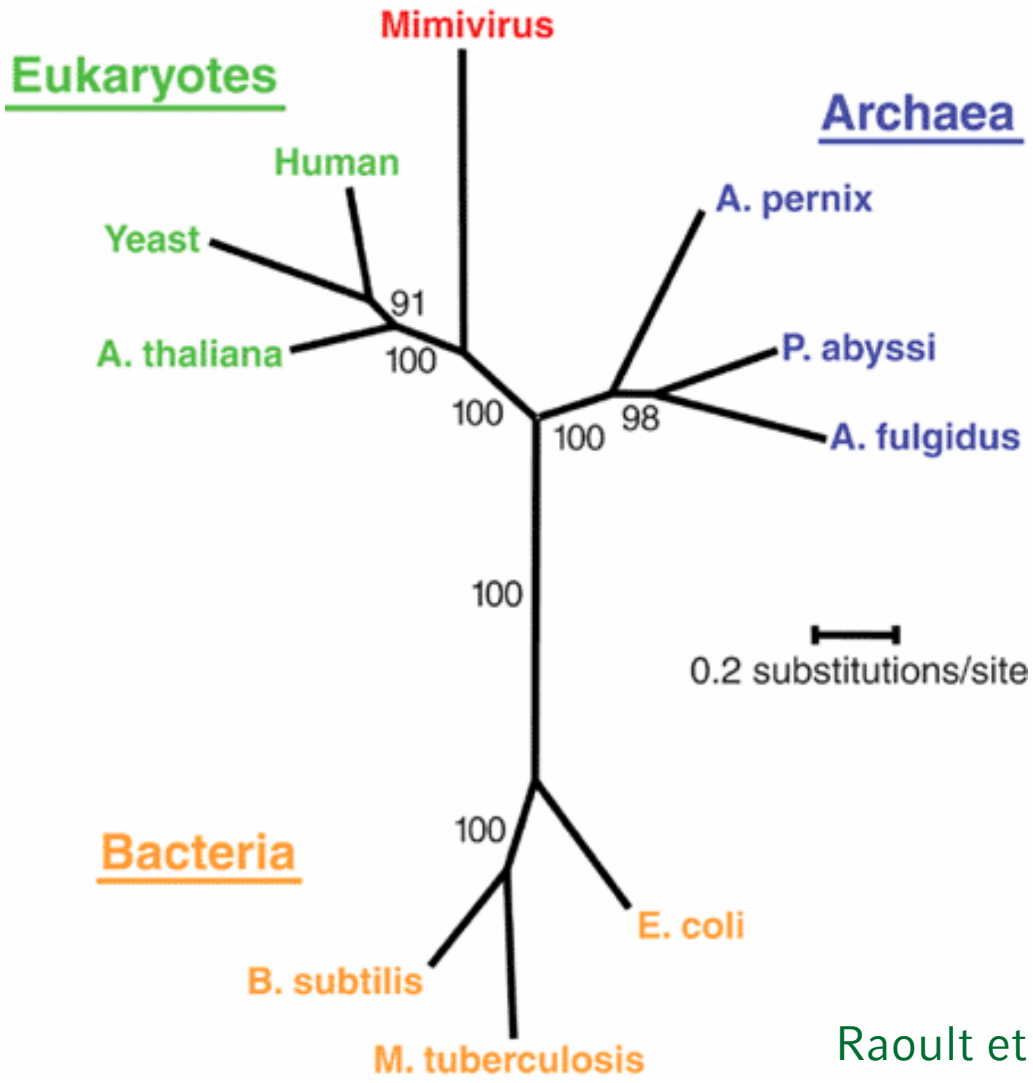
- insb. Proteine zur Proteinsynthese
  - 4 tRNA-Synthase Gene
  - 4 weitere Gene zur Proteinsynthese
  - 6 tRNA Gene
- Diese Proteine waren bislang „lebenden“ Organismen vorbehalten!

The 1.2-Megabase Genome Sequence of Mimivirus,
D. Raoult, S. Audic, C. Robert, C. Abergel, P. Renesto, H. Ogata,
B. La Scola, M. Suzan, J.-M. Claverie,
Science (19 NOV 2004) **306**: 1344-1350.

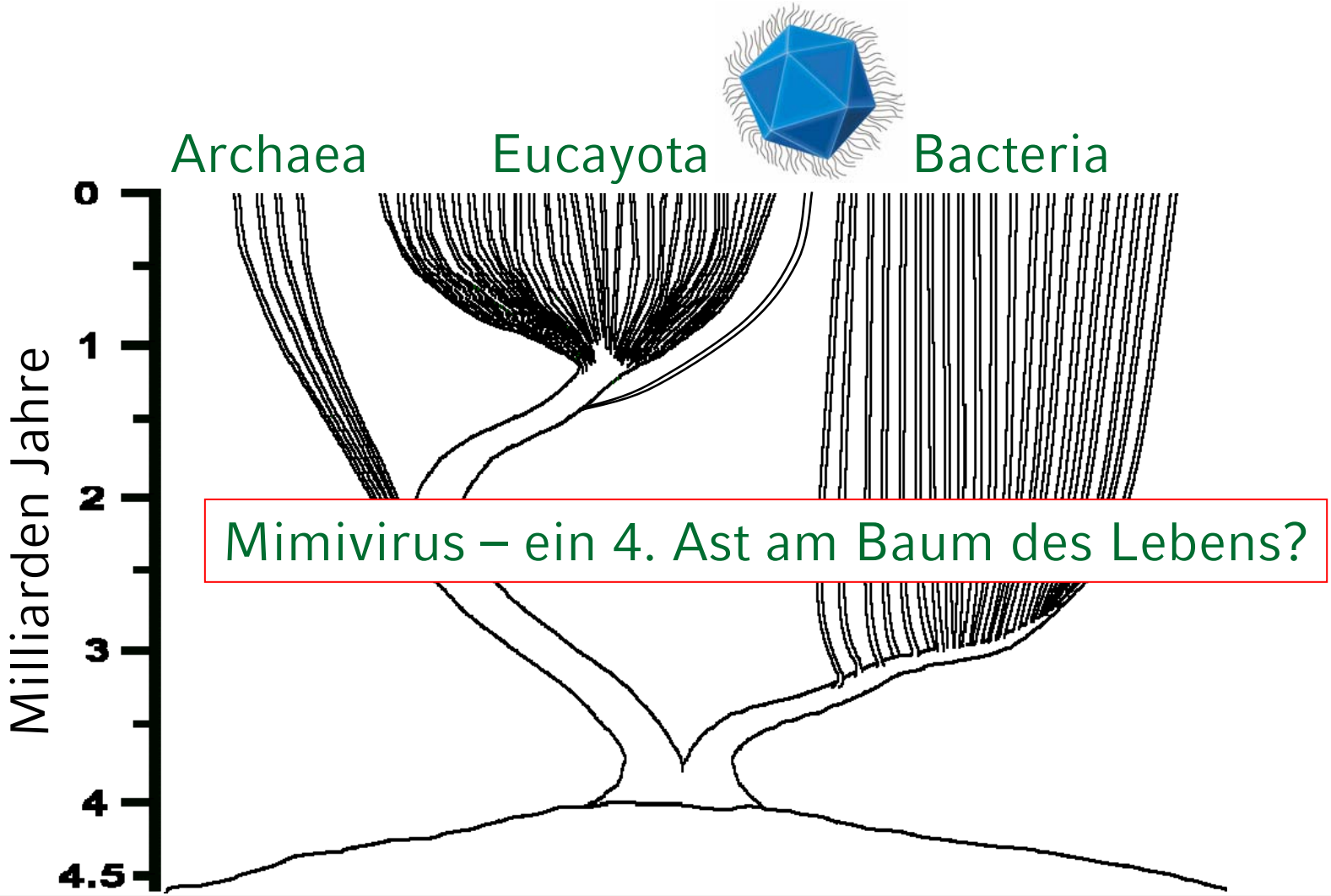


Der Vergleich von „homologen“ Proteinen erlaubt die Rekonstruktion der Entwicklungsgeschichte der Organismen

- Auf der Basis von 7 ubiquitären Proteinen ....
- ... wurde ein „phylogenetischer Baum“ berechnet



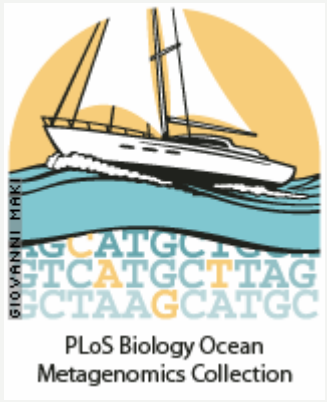
Raoult et al., Science, 2004.



Mimivirus – ein 4. Ast am Baum des Lebens?

Die Jagd ist offen ...

„Metagenomics“
Craig Venter *et al.*



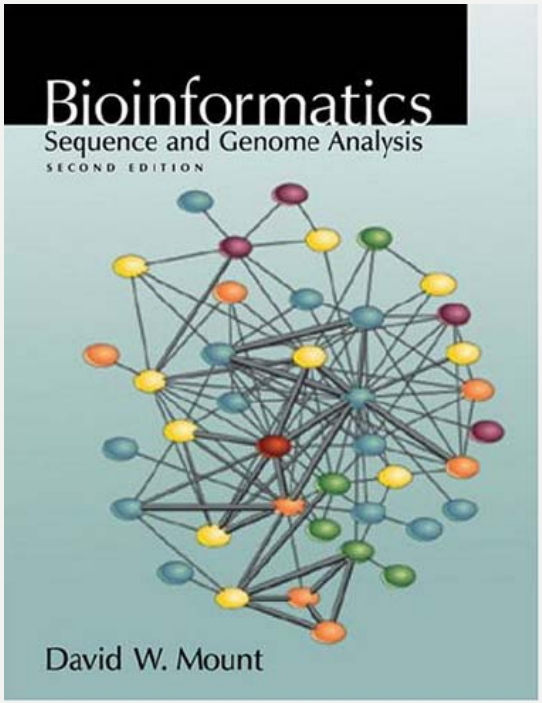
Cousins von Mimivirus wurden inzwischen in der Sargasso See entdeckt

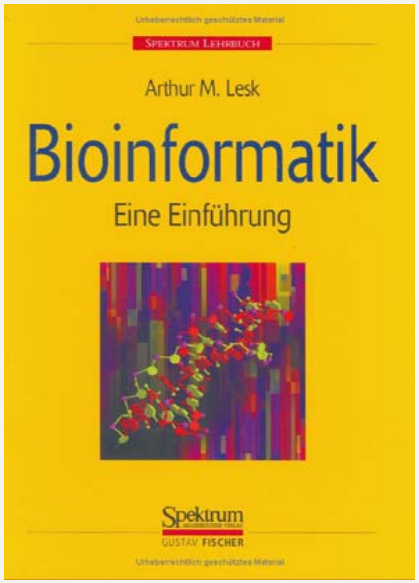
LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

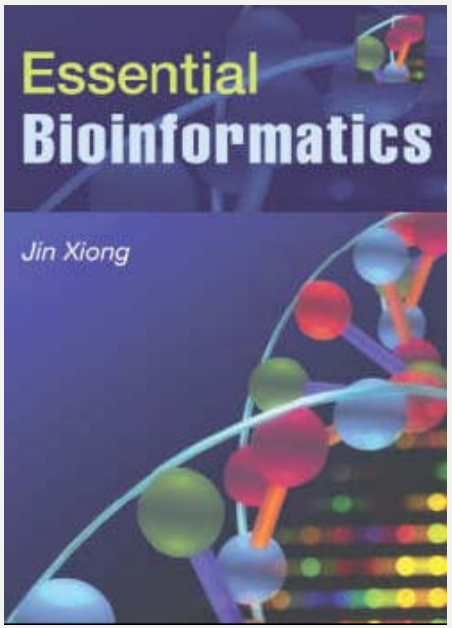
Vielen Dank für Ihr Interesse

Fragen und Anmerkungen gerne an
karsten.suhre@lmu.de
Tel.: 3187-2627

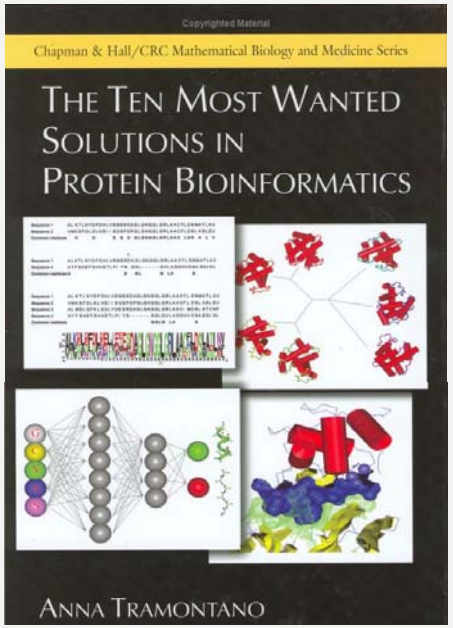
Literatur

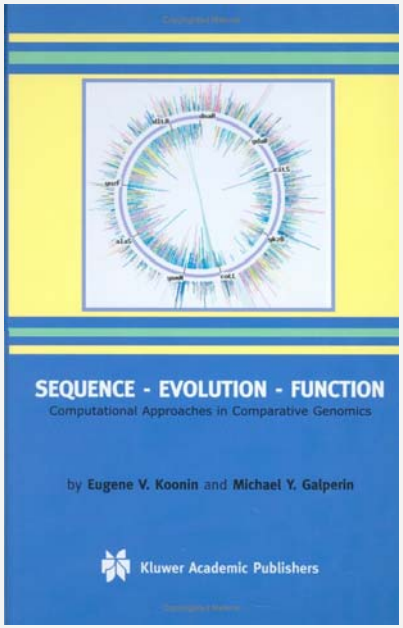

Bioinformatics
Sequence and Genome Analysis
SECOND EDITION
David W. Mount

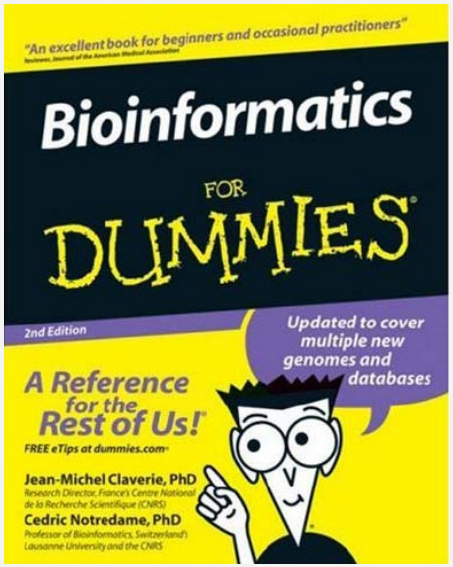

Arthur M. Lesk
Bioinformatik
Eine Einführung
Spektrum


Essential
Bioinformatics
Jin Xiong

Literatur


Copyrighted Material
Chapman & Hall/CRC Mathematical Biology and Medicine Series
THE TEN MOST WANTED SOLUTIONS IN PROTEIN BIOINFORMATICS
ANNA TRAMONTANO


SEQUENCE - EVOLUTION - FUNCTION
Computational Approaches in Comparative Genomics
by Eugene V. Koonin and Michael Y. Galperin
Kluwer Academic Publishers


"An excellent book for beginners and occasional practitioners"
Bioinformatics
FOR
DUMMIES
2nd Edition
Updated to cover multiple new genomes and databases
A Reference for the Rest of Us!
FREE eTips at dummies.com
Jean-Michel Claverie, PhD
Cedric Notredame, PhD

Literatur

- D. W. Mount, *Bioinformatics: Sequence and Genome Analysis*, CSHL Press.
- A.M. Lesk, *Bioinformatik, Eine Einführung*, Spektrum.
- J. Xiong, *Essential Bioinformatics*, Cambridge University Press.
- A. Tramontano, The Ten Most Wanted Solutions in Protein Bioinformatics, Chapman&Hall/CRC.
- E.V. Koonin & M.Y. Galperin, *Sequence-Evolution-Function*, Springer.
- J.M. Claverie & C. Notredame, *Bioinformatik for Dummies*, Wiley.

- NCBI Bookshelf: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=Books

Copyright Hinweis

Diese Präsentation ist als Unterrichtsmaterial für die Vorlesung ‘Seniorenstudium: Mit dem Computer auf der Spur von Mimi, dem Riesenvirus’ bestimmt. Soweit Darstellungen aus öffentlich zugänglichen Quellen (Literatur, WWW) verwendet werden, sind diese Quellen angegeben. Das Lehrmaterial ist ausschließlich für die Studenten dieser Vorlesung gedacht. Vervielfältigung und Veröffentlichung sind aus urheberrechtlichen Gründen untersagt.